NAVAL POSTGRADUATE SCHOOL
Monterey, California

AD-A267 442

THESIS

PERFORMANCE OF CODED COHERENT FSK LIGHTWAVE
SYSTEM WITH NONCOHERENT DETECTION

by

Thomas Elliott Neely

March, 1993

Thesis Advisor: Tri T. Ha
Co-Advisor: Daniel J. Collins

Approved for public release; distribution is unlimited

93-16860

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION Naval Postgraduate School	6b. OFFICE SYMBOL (If applicable) 34	7a. NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000		7b. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)		10. SOURCE OF FUNDING NUMBERS

Program Element No.	Project No.	Task No.	Work Unit Accession Number

11. TITLE (Include Security Classification)
PERFORMANCE OF CODED COHERENT FSK LIGHTWAVE SYSTEM WITH NONCOHERENT DETECTION

12. PERSONAL AUTHOR(S) Thomas Elliott Neely

13a. TYPE OF REPORT Master's Thesis	13b. TIME COVERED From To	14. DATE OF REPORT (year, month, day) March 1993	15. PAGE COUNT 76

16. SUPPLEMENTARY NOTATION
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

17. COSATI CODES			18. SUBJECT TERMS (continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUBGROUP	Optical Communications, Laser Phase Noise, Coding

19. ABSTRACT (continue on reverse if necessary and identify by block number)

The original coherent lightwave systems were expected to offer significant performance gains relative to standard direct detection systems. This expectation has not been realized due to the effects of laser phase noise. The laser phase noise process results in the integration of a random variable that transitions over the integration period from a Gaussian distribution to a uniform distribution. The use of convolutional coding effectively replaces a single bit time, with its mostly noncoherently integrating latter portion, by several more coherently integrating bits. This primary bit-time effect comes in addition to the normal coding effect of efficiently trading bandwidth for error performance. The improvement in performance brought about from coding may enable coherent systems to live up to previous expectations.

The contributions of this thesis include the visualization of the phase noise process, the efficient computation of the laser phase noise power factor probability density function, and the computation of performance curves for uncoded and coded systems. Additional sections on coherent lightwave systems and coding provide tutorial information. A potential military application is discussed, along with practical implementation issues.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS REPORT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Ha, Tri T. and Collins, Daniel J.	22b. TELEPHONE (Include Area code) (408) 656-2788; (408) 656-2491	22c. OFFICE SYMBOL EC/HA ; AA/CO

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted   SECURITY CLASSIFICATION OF THIS PAGE
All other editions are obsolete

UNCLASSIFIED

Approved for public release; distribution is unlimited

PERFORMANCE OF CODED COHERENT FSK LIGHTWAVE SYSTEM WITH NONCOHERENT DETECTION

by

Thomas Elliott Neely
Lieutenant, United States Navy
B.S., Ohio State University, 1983

Submitted in partial fulfillment of the requirements for the degrees of

MASTER OF SCIENCE IN AERONAUTICAL ENGINEERING

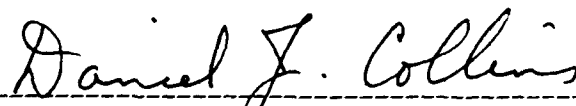**MASTER OF SCIENCE IN ELECTRICAL ENGINEERING**

from the

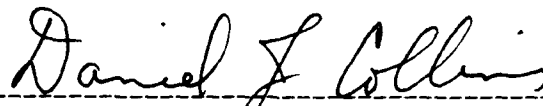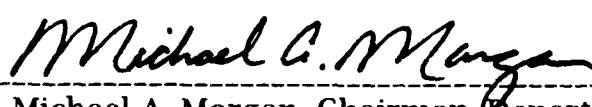**NAVAL POSTGRADUATE SCHOOL**
March 1993

Author: _____
Thomas Elliott Neely

Approved By: _____
Tri T. Ha, Thesis Advisor

_____
Daniel J. Collins, Thesis Co-Advisor

_____
Daniel J. Collins, Chairman, Department of
Aeronautical and Astronautical Engineering

_____
Michael A. Morgan, Chairman, Department of
Electrical and Computer Engineering

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

ii

ABSTRACT

The original coherent lightwave systems were expected to offer significant performance gains relative to standard direct detection systems. This expectation has not been realized due to the effects of laser phase noise. The laser phase noise process results in the integration of a random variable that transitions over the integration period from a Gaussian distribution to a uniform distribution. The use of convolutional coding effectively replaces a single bit time, with its mostly noncoherently integrating latter portion, by several more coherently integrating bits. This primary bit-time effect comes in addition to the normal coding effect of efficiently trading bandwidth for error performance. The improvement in performance brought about from coding may enable coherent systems to live up to previous expectations.

The contributions of this thesis include the visualization of the phase noise process, the efficient computation of the laser phase noise power factor probability density function, and the computation of performance curves for uncoded and coded systems. Additional sections on coherent lightwave systems and coding provide tutorial information. A potential military application is discussed, along with practical implementation issues.

TABLE OF CONTENTS

I. INTRODUCTION

Optical communications links have become the preferred channels for high data rate applications for many reasons. A single fiber can support the bandwidth of hundreds of coaxial cables; fibers are immune from radio frequency interference; damage to fiber cables is easily located; fiber is lower in weight than metal cables. As fiber cables have become more rugged, their use in military systems has become more practical. Fiber is now being proposed for guided weapon data links, battlefield communication networks, and sensor systems.

The current commercial state of the art is direct detection. Direct detection lightwave systems are similar in concept to A.M. radio, in that the process of detecting the signal makes no use of the signal phase. In essence, the signals consist of pulses of noise energy, with a bit "1" represented by a pulse of energy, and a bit "0" represented by no signal. The preferred channel of propagation is single mode fiber. Fibers are cylindrical waveguides subject to the same physical laws as microwave waveguides. The term "single-mode" refers to the support of only one propagating mode. This eliminates dispersion due to intermodal mixing. Fibers are subject to attenuation, which is logarithmic with distance. Fibers also suffer from chromatic dispersion, where the spectral components of the source's non zero linewidth travel at different velocities through the waveguide material. Chromatic dispersion effects are reduced through the use of laser diodes with small linewidths.

Attenuation is reduced through the formulation of fiber materials and strict elimination of water and OH^- molecules from the fiber during manufacturing.

Improved performance is theoretically possible using coherent systems. In a coherent system receiver, the source light signal is optically mixed with a local oscillator optical signal, resulting in an electronic intermediate frequency signal. The electronic IF signal is processed the same way as in a microwave coherent communication system. The source and local oscillator lasers must be carefully controlled to maintain an IF signal compatible with the demodulator. This makes coherent systems more complicated than direct detection systems. Coherent receivers show significant sensitivity improvements (up to 20 dB) as compared to standard direct detection systems.

A. THE PROBLEM, HISTORY, AND THIS THESIS

The significant barrier to the use of coherent systems is the presence of phase noise in the source laser and in the receiver local oscillator laser. The phase noise results in source linewidth broadening. The nature of phase noise is such that its effects *decrease* with increasing data rates. The commercial implementation of coherent systems awaits the development of laser sources with small linewidth and rapid modulation capability.

The effects of phase noise became known after early experiments in coherent systems resulted in dismal performance. One of the first comprehensive papers on the nature of phase noise was published by

2

Salz in 1985 (Salz, 1985). The significant obstacle to quantifying the effects of phase noise is the determination of the probability density function for the phase noise process. Approximations to the density were put forth in a series of papers by Foschini, Greenstein and Vannucci, (Foschini, Greenstein, and Vannucci, 1988; and Foschini and Vannucci, 1988). Azizoglu and Humblet (Azizoglu and Humblet, 1991) refined the earlier treatment, and demonstrated how multisampling receivers can combat the effects of phase noise. Since multisampling is equivalent to diversity, and since diversity is equivalent to simple repetition coding, it makes sense that the application of more efficient forms of coding can provide greater benefits than diversity.

This thesis is about the application of convolutional coding to binary frequency shift keying (2FSK) coherent lightwave systems with noncoherent detection. It is shown that coding offers significant improvements in error performance for a given data rate.

B.   CHOICES, TRADEOFFS, AND COSTS

The choice of noncoherent detection is driven by the more relaxed linewidth requirements for this type of system. With the choice of noncoherent detection, the possible modulation formats are on-off keying (OOK), and frequency shift keying (FSK). The error performance of these two systems referenced to average bit energy to noise ratios are the same, but the peak power required in the FSK system to achieve equal average power is 3 dB less than that for OOK. The type of coding considered is hard decision convolutional codes. While soft decision

decoding offers several dB improvement, it is more complicated and requires a decoder that is more difficult to implement at high signal rates.

The phase noise effect decreases with data rate, and it will be shown that for uncoded systems, the drop in bit energy from an increase in data rate is generally more than offset by the drop in phase noise effect. Practical systems are limited in signal rate; it is not always possible to increase coding gain without encountering this signal rate limit. This means that there will be tradeoffs between increasing data rate and employing coding. The requirements placed on the system must be stated in terms of required data rate, limits on signal rate, and error performance. In many circumstances, employing coding will be beneficial, in others, it may not.

The tradeoff options mount when the ability to frequency multiplex coherent systems within common channels is considered. Now, several lower data rate coded channels may be multiplexed, providing the data rate of a single uncoded channel, at lower total power (but greater complexity). The error performance curves presented in this thesis can be utilized to conduct tradeoff studies between single channel and multiplexed systems.

The cost of implementing coding may be very small. For convolutional coding, the encoders are very simple. The decoders are somewhat more complicated. The Viterbi decoders are well suited to implementation as application specific integrated circuits (ASICs). Several companies advertise decoders for low data rates (10 Mbps) for

4

$25 per chip. The fastest available Viterbi decoder is a hard decision system capable of operating at signal rates of 200 Mbps (NASA, 1993). The capabilities of decoders will increase as logic speeds increase and parallel architectures are exploited.

C.   THESIS OUTLINE

The outline of the thesis is as follows. The first section provides background information of a tutorial nature on optical communication systems, and the phase noise process. The next section develops the effects of phase noise on uncoded 2FSK systems with noncoherent detection. New contributions in this area include the visualization of the phase noise process, and the efficient computation of the probability density function from its approximate moments. The next section begins with tutorial material on convolutional coding, and then develops the error performance for coded systems using several simple codes. The concluding section considers a hypothetical coded coherent system and compares it to a conventional direct detection implementation. An appendix contains computer codes of the probability density function (pdf) computation and code transfer function determination for the Mathematica computer algebra system.

II. OPTICAL COMMUNICATION SYSTEMS

Lightwave systems may be broadly classified as noncoherent systems or coherent systems.

A. NONCOHERENT SYSTEMS

In a noncoherent system (Figure 2.1), a source laser is intensity modulated, the resulting optical signal is sent over multimode or single mode fiber cables, or propagated through free space. Fiber optic cables are cylindrical waveguides, and thus light propagates under the same physical laws as apply to microwave waveguides, i.e., the waveguide can propagate either single or multiple modes. Multiple mode fibers are affected by the interactions between the modes, causing modal dispersion, which widens the spectrum of transmitted light. The longer the cable, the more severe the modal dispersion becomes, leading to performance degradation. Obviously, a single mode fiber does not suffer from modal dispersion. Another form of dispersion that all fibers suffer from is chromatic dispersion, where different frequencies travel at different velocities due to a non constant index of refraction. The solution to this problem is to use a source with as small a linewidth as possible.
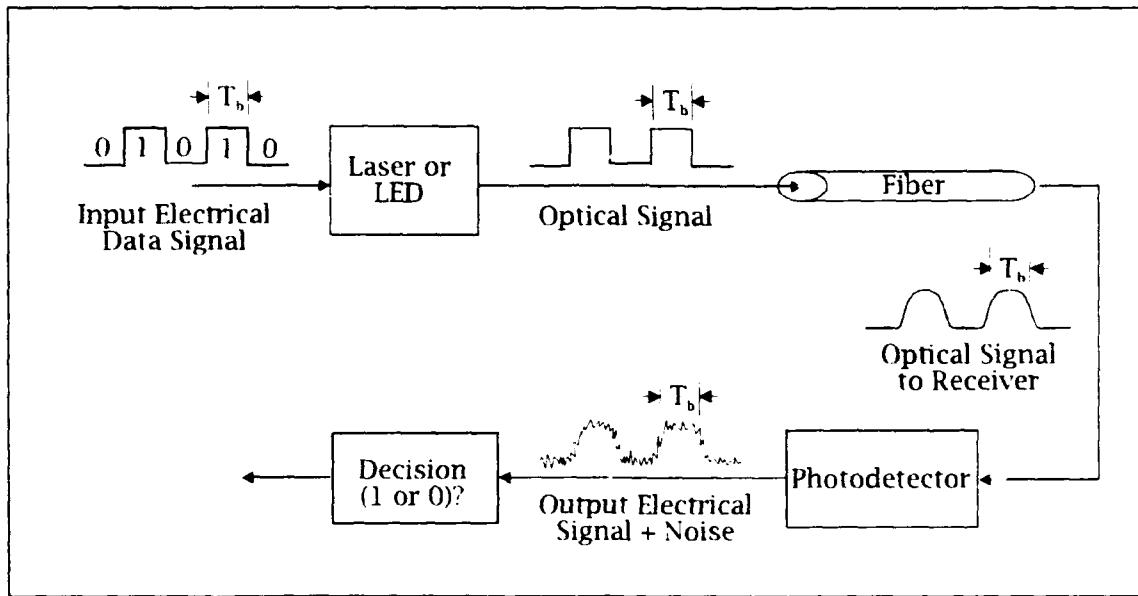
Figure 2.1   Noncoherent Lightwave System

When employing wide linewidth sources (like LED's) over multimode
fibers, the maximum link distance is on the order of 10 Km (Keiser,
1991). At the receiver, the intensity modulated signal power is directly
detected by a photodiode, generating a voltage proportional to received
optical power plus noise. The similarity between noncoherent lightwave
systems and a crystal A.M. radio are now clear. Advantages of
noncoherent systems include simplicity, and the ability to use relatively
noncoherent light sources such as light emitting diodes, as opposed to
more coherent, more complicated, and more costly laser diodes. The
disadvantages of noncoherent systems are difficulties in multiplexing
multiple signals on multiple lightwave frequencies, and inferior error
and inferior repeaterless range compared to coherent systems.

7

B.   COHERENT SYSTEMS

Coherent lightwave systems (Figure 2.2) employ a spectrally very pure source laser, and single mode propagation channels. The spectral widening through a multimode fiber would be intolerable in a coherent system, so current systems employ single mode fibers. Free space, low dispersion propagation is also possible. At the receiver, the incoming laser light is first mixed with a local oscillator laser (Figure 2.3). When the mixing is to baseband, the process is termed homodyne detection. When the mixing is to an intermediate frequency, it is termed heterodyne detection. The detection and decision process is now carried out electronically, exactly as in a coherent microwave communication system. Coherent systems show up to 20 dB better receiver sensitivity compared to noncoherent systems (Keiser, 1991), and also a high degree of frequency selectivity. The repeaterless range of coherent systems is on the order of 100 Km. Figure 2.4 shows how a large number of channels (assuming 10 GHz channel spacing) can theoretically be fit into the windows of low fiber attenuation available with current optical fibers. Further information on fiber optic systems can be found in (Keiser, 1991).

The nature of coherent systems demands close control over the frequencies of source and local lasers, which adds complexity to coherent systems. The frequency of current laser diodes is affected by the laser temperature, and the drive current. Modulating the drive current is one method to impart frequency modulation on the source laser. Typical sensitivities of a laser are 10-20 GHz/°C temperature and

1-5 GHz/mA drive current. Since the intermediate frequency is usually
in the MHz range, slight deviations in temperature or drive current in
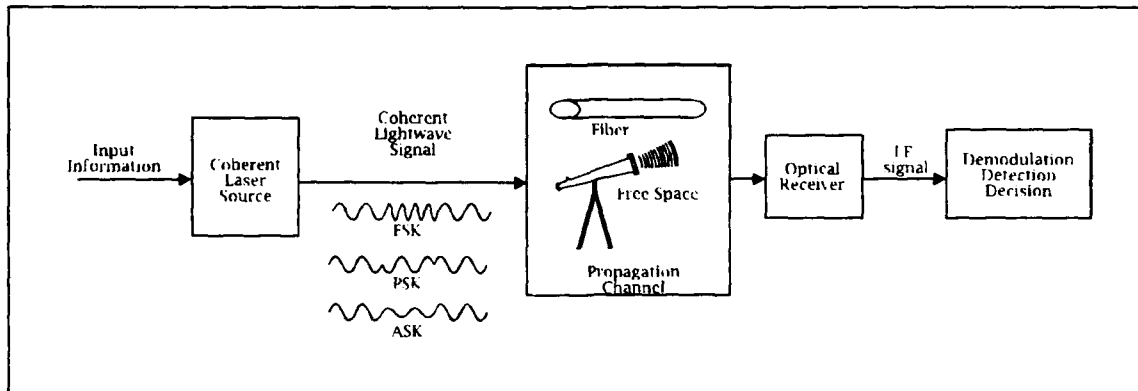source or local lasers can cause trouble.

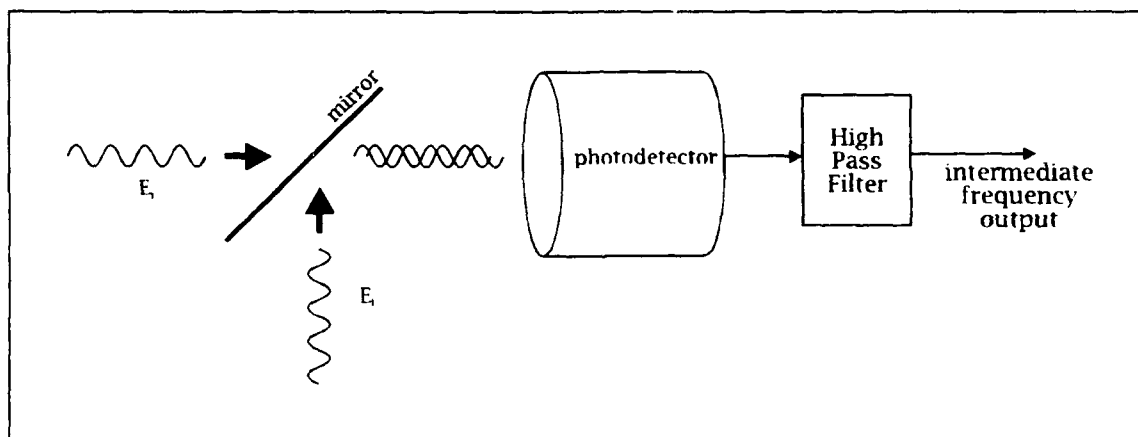Figure 2.2   Coherent Heterodyne Lightwave System

Figure 2.3   Heterodyne Optical Receiver

C.   OPTICAL RECEIVERS

A heterodyne optical receiver is shown in Figure 2.3. In this
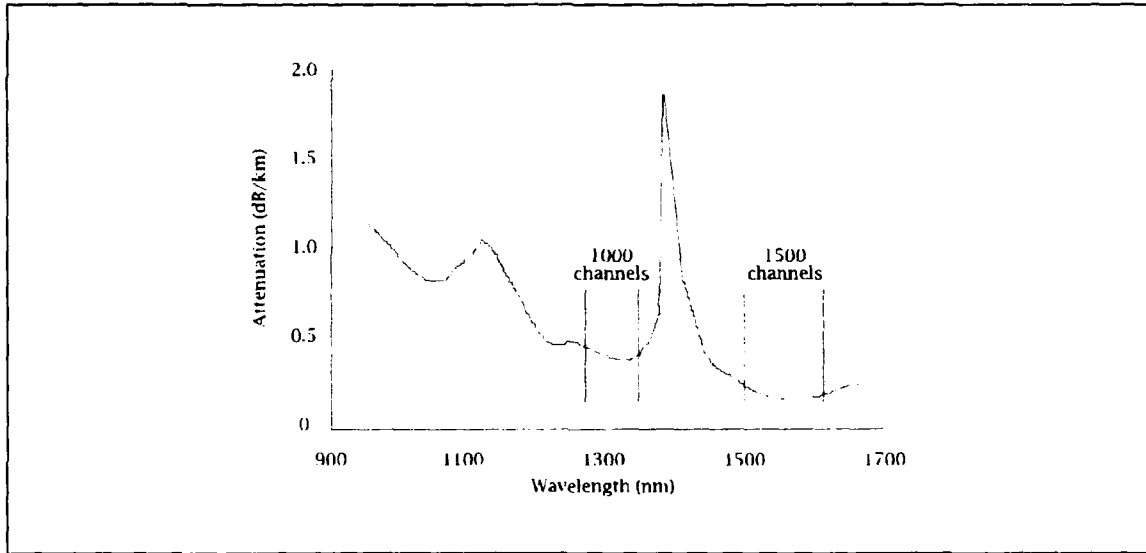receiver, the incoming source signal is projected onto the photodetector

Figure 2.4   Fiber channel capacity assuming 10 GHz channel spacing

surface. A local oscillator laser is optically mixed via a mirror onto the

photodetector.   Denote the source signal field by

$$E_r = A_r(t)\cos[2\pi f_r t + \theta_r(t)] \tag{2.1}$$

where f_r is the carrier frequency of the source laser, $\theta_r(t)$ is the source

laser phase, and $A_r(t)$ is the envelope. The polarization angle between

the two beams is $\phi(t)$. Likewise, the local oscillator laser field is

$$E_l = A_l \cos[2\pi f_l t + \theta_l(t)] \tag{2.2}$$

where f_l is the local frequency, $\theta_l(t)$ is the local phase, and A_l is the

(constant) local envelope. The photodetector responds to the combined

fields of $\{r(t) + l(t)\}$ with intensity

$$I_c(t) = \frac{1}{2}A_r^2(t) + \frac{1}{2}A_l^2 + A_r(t)A_l \cos[2\pi(f_r - f_l) + \theta_r(t) + \theta_{l(t)}]\cos\phi(t) \tag{2.3}$$

The output of the photodetector therefore has an intermediate frequency term at $f_r - f_l$. The term at $f_r + f_l$ has been discarded in (2.3).

The intermediate frequency power at the photodetector will be

$$P_c(t) = 2\sqrt{P_r P_l}\cos\left[2\pi(f_r - f_l)t + \theta_r(t) - \theta_l(t)\right]\cos\phi(t) \tag{2.4}$$

A noncoherent, direct detection system responds only to the power of the incoming signal, with an intensity of

$$I_{dd}(t) = \frac{1}{2}A_r^2(t)\left[1 + \cos(4\pi f_r t + 2\theta_r(t))\right] \tag{2.5}$$

The double frequency term will normally be outside the photodetector's response. The power at the photodetector will be

$$P_{dd}(t) = P_r \tag{2.6}$$

To determine why coherent systems offer superior performance, the quantum performance of a photodetector must be considered. A photodetector requires a certain number N_p of photons to fall on the detector to produce a single electron. For an ideal detector, $N_p = 1$. The reciprocal of N_p is called the quantum efficiency of the detector, and is denoted by η. The current produced by M photons/sec is $i = M\eta q$, where q is the electron charge. Each photon of light has energy hf where h is Planck's constant, and f is the light frequency. Therefore, the incident power $P = Mfh$ and $i = \eta q P/hf$. As the number of photons falling on the photodetector becomes small, the discrete nature of the photon to electron conversion events leads to a fluctuation in the output current that becomes more severe as the number of photons falling on the

11

detector decreases. This is a noise process, and is called shot noise. For a signal bandwidth b, the shot noise process has power (DeLange, 1968)

$$N_s = \frac{2q^2 \eta b P}{hf} R \qquad (2.7)$$

The signal power is $i^2 R$, so the signal to shot noise ratio is

$$\frac{S}{N} = \frac{i^2 R}{N_s} = \frac{\eta P}{2hfb} \qquad (2.8)$$

The above analysis is simplified, in that noise components resulting from dark currents and thermal noise are not included. A more complete analysis may be found in (Keiser, 1991).

Comparing (2.4), (2.6), and (2.8) reveals the superiority of coherent systems. Since the local oscillator power carries the signal modulation, the photodetector may work at much higher signal to noise ratios. Currently, coherent receivers achieve at least 10 dB better sensitivity for a given SNR than noncoherent receivers.

D. DEMODULATORS

Coherent heterodyne lightwave systems are further classified as to the processing at IF. In coherent demodulation, the phase of the local oscillator (within the detector, not the local oscillator laser) is matched precisely to that of the IF signal via a phase locking technique. A coherent 2FSK system is depicted in Figure 2.5. In this figure, the incoming and local oscillators have equal phase $\theta(t)$, and the input signal is contaminated by noise $n(t)$. The coherent system is the optimum approach when the phase of the source and local lasers is known and

12

stable. Unfortunately, stable phase is difficult and costly to achieve.
The alternative, which costs 3 dB in theoretical signal to noise ratio, is to
use noncoherent demodulation (Figure 2.6). In this scheme, the phase of
the IF signal is irrelevant, since both inphase and quadrature
components of the detection process are squared and added.
Fluctuation in the source laser and laser local oscillator phase (phase
noise) causes fluctuation in the IF phase and still degrades the
performance, but the phase noise does not of itself complicate the
detection process. The noncoherent 2FSK system depicted in Figure 2.6
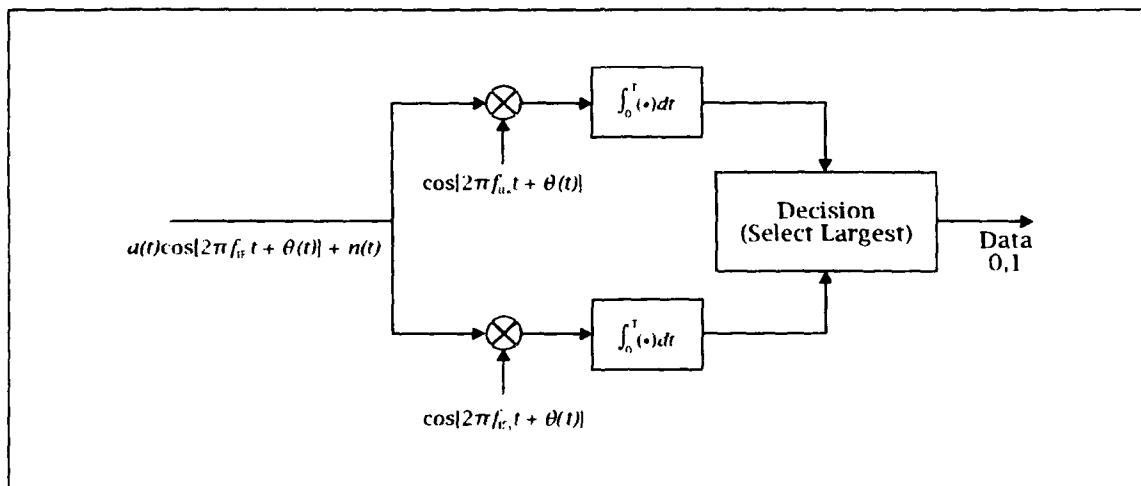is the system studied in this thesis.
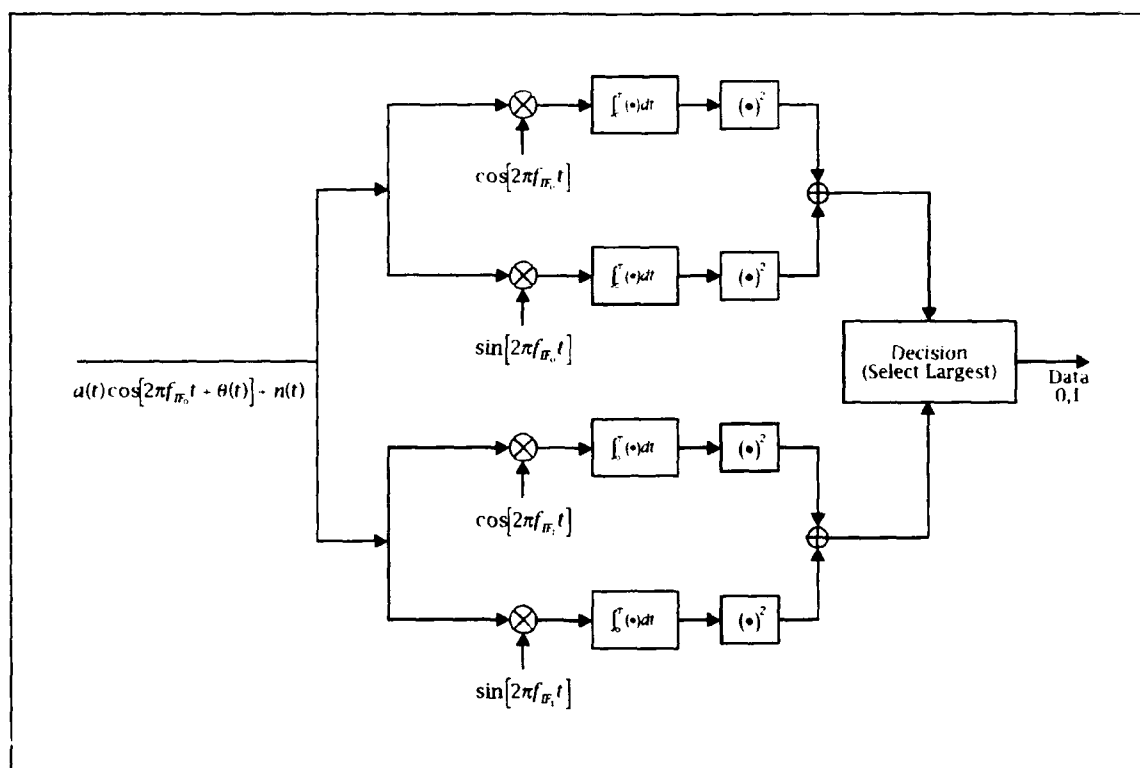

Figure 2.5   Coherent 2FSK Demodulation

Figure 2.6   Noncoherent 2FSK demodulator

III. LASER PHASE NOISE

For a variety of practical reasons, the preferred devices for commercial and military applications of fiber optic communications are semiconductor injection lasers (laser diodes). It is the relative spectral purity of the laser diode that permits the employment of coherent optical modulation schemes. The spectral purity is not absolute, however. These devices are generally affected by phase noise, which broadens the spectrum of the laser and can severely degrade performance.

Earlier it was stated that coherent receivers enjoy a 10 dB sensitivity advantage over direct detection systems. In practice, this advantage is only realizable when the effects of phase noise are controlled.

Laser phase noise is caused by random spontaneous emissions within the lasing medium. Each emission event causes a random jump in the phase. This is a random walk process, as time progresses, the value of the phase will wander away from where it started. The mean squared deviation of the phase grows linearly with time. It has been shown (Salz, 1985) that the common Lorentzian shape of the laser spectrum is due to this phase noise process.

Since the mean time between the random jumps is infinitesimal, the phase noise process becomes in the limit a Wiener process characterized by a zero mean white Gaussian noise $\mu(t)$ with some two sided power spectral density N_0

$$\theta(t) = 2\pi \int_0^t \mu(\tau)d\tau \qquad (3.1)$$

with

$$\begin{aligned}
E\{\theta^2(t)\} &= E\left\{4\pi^2 \int_0^t \mu(\tau_1)d\tau_1 \int_0^t \mu(\tau_2)d\tau_2\right\} \\
&= 4\pi^2 \int_0^t \int_0^t N_0\delta(\tau_1 - \tau_1)d\tau_1 d\tau_2 \qquad (3.2)\\
&= 4\pi^2 N_0 t
\end{aligned}$$

To determine a value for N_0, consider the laser emission process

$$s(t) = \cos\left[2\pi f_0 t + \theta(t) + \phi\right] \qquad (3.3)$$

where ϕ is a uniformly distributed random variable.

The power spectral density of $s(t)$ is (Salz, 1985)

$$G(f) = \frac{1}{4\pi^2 N_0}\left\{\frac{1}{1+\left(\dfrac{f+f_0}{\pi N_0}\right)^2} + \frac{1}{1+\left(\dfrac{f-f_0}{\pi N_0}\right)^2}\right\} \qquad (3.4)$$

This is the Lorentzian line shape shown in Figure 3.1. Denoting the half power linewidth as β, equation (3.4) may be solved for

$$N_0 = \frac{\beta}{2\pi} \qquad (3.5)$$

The random phase variable $\theta(t)$ is therefore a zero mean white Gaussian process with PSD from (3.2) of $2\pi\beta t$.
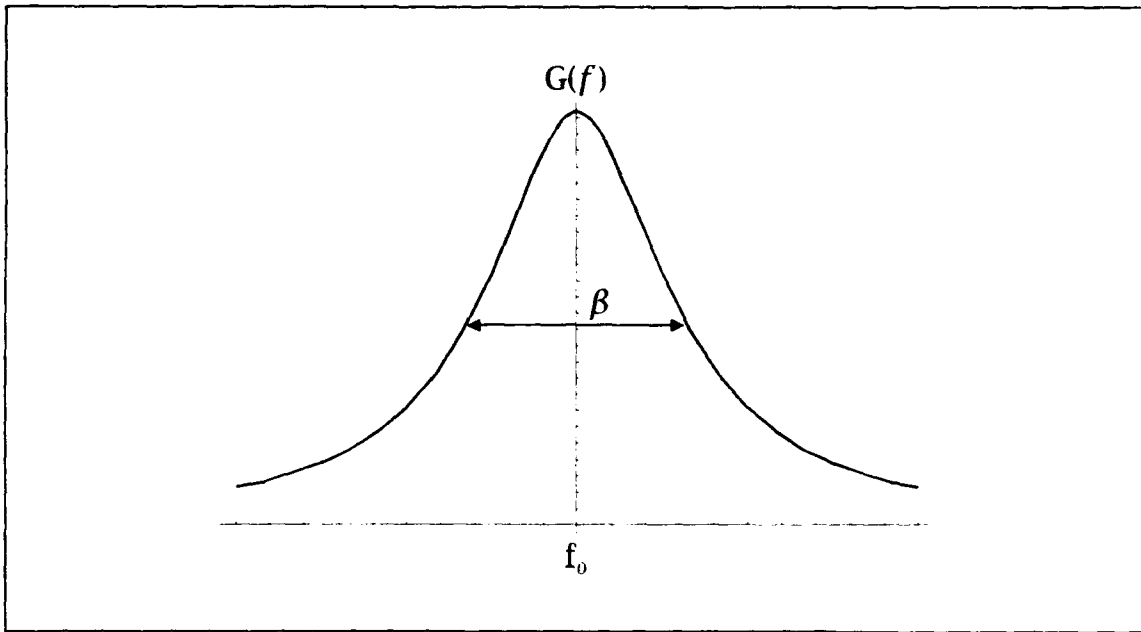
16

Figure 3.1   Laser Spectrum

IV. UNCODED SYSTEM PERFORMANCE

The coherent system with noncoherent demodulation of Figure 4.1 will be considered to determine the deleterious effects of phase noise.

A.   NONCOHERENT 2FSK ERROR PERFORMANCE FORMULATION

The input to the demodulator is the IF output from the optical heterodyne receiver

$$r(t) = A\cos[2\pi f_0 t + \theta(t)] + n(t) \tag{4.1}$$

It is assumed that a data 0 is transmitted. The signal is contaminated by additive white Gaussian noise with variance $N_0/2$. From (3.2) and (3.5), the phase noise $\theta(t)$ is white Gaussian with variance $2\pi\beta t$. It is assumed that the frequency spread between f_0 and f_1 is large enough to assure orthogonality.

The decision variables are

$$Y_0 = \left| \frac{A}{2} \int_0^T \exp[j\theta(t)]dt + n_{i_0} + jn_{q_0} \right|^2$$

$$Y_1 = \left| n_{i_1} + jn_{q_1} \right|^2 \tag{4.2}$$

where the n are independent identically distributed white Gaussian random variables with variance $N_0 T/4$. If the phase noise integral term is defined as

$$Y^2 = \left| \frac{A}{2} \int_0^T \exp[j\theta(t)]dt \right|^2 \tag{4.3}$$
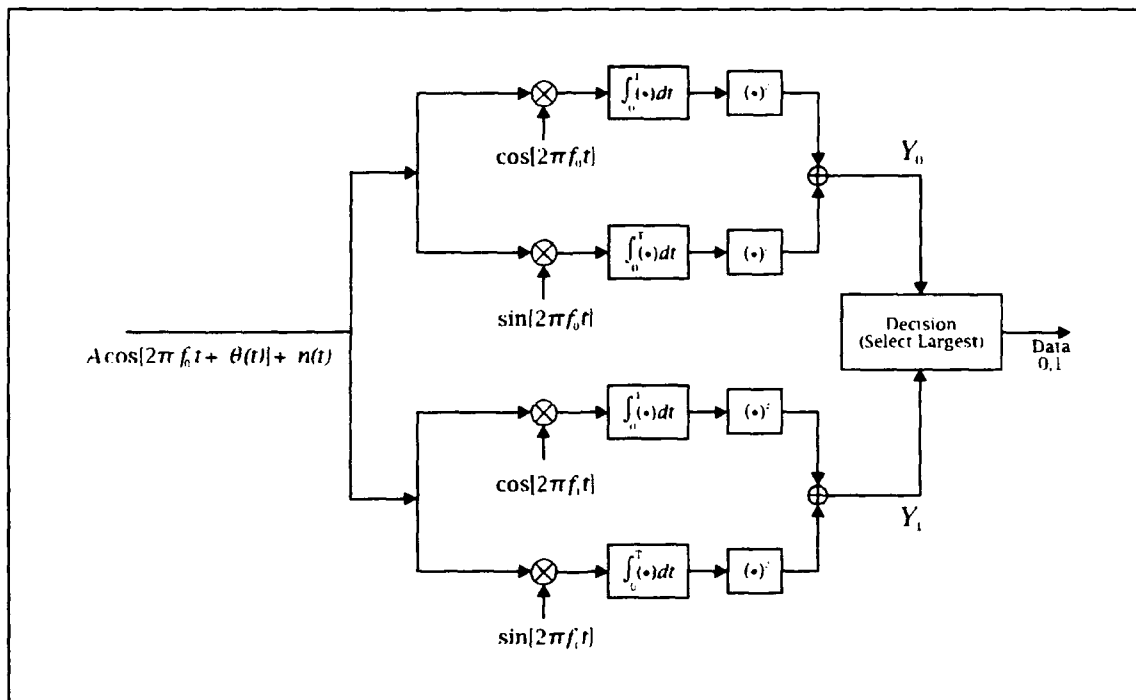
18

Figure 4.1   Noncoherent Demodulator

then the probability of error conditioned on Y yields the standard noncoherent detection problem. The density for Y_0 conditioned on Y is the noncentral χ^2 distribution

$$f_{Y_0|Y}(y_0) = \frac{1}{2\sigma^2} \exp\left[\frac{-\left(y_0 + Y^2\right)}{2\sigma^2}\right] I_0\left[\frac{Y\sqrt{y_0}}{\sigma^2}\right] \qquad (4.4)$$

where I_0 is the modified Bessel function of zero order, and $\sigma^2 = N_0 T/4$. The density for Y_1 is the exponential density

$$f_{Y_1}(y_1) = \frac{1}{2\sigma^2} \exp\left[\frac{-y_1}{2\sigma^2}\right] \qquad (4.5)$$

with $\sigma^2 = N_0 T/4$. The error probability conditioned on Y becomes

$$\Pr\{Y_1 > Y_0 | Y\} = \frac{1}{2} \exp\left[\frac{-Y^2}{4\sigma^2}\right]$$

$$= \frac{1}{2} \exp\left[\frac{-Y^2}{N_0 T}\right]$$

(4.6)

When the signal is not corrupted by phase noise, the bit error probability is

$$P_b = \frac{1}{2} \exp\left[-\frac{1}{2}\frac{\mathcal{E}_b}{N_0}\right]$$

$$= \frac{1}{2} \exp\left[\frac{-A^2 T}{4 N_0}\right]$$

(4.7)

where $\mathcal{E}_b/N_0$ is the signal energy to noise ratio.

The expression for Y may be rewritten

$$Y^2 = \left|\frac{A}{2} \int_0^T \exp[j\theta(t)]dt\right|^2$$

$$= \frac{A^2 T^2}{4} \left|\int_0^1 \exp[j\theta(xT)]dx\right|^2$$

(4.8)

$$= \frac{A^2 T^2}{4} \left|\int_0^1 \exp[j\sqrt{\gamma}\psi(x)]dx\right|^2$$

by making the change of variable $t = xT$. The random variable $\psi(x)$ has $E\{\psi^2(x)\} = x$, and $\gamma = 2\pi\beta T$, where β is the combined 3 dB linewidths of the source and local lasers and T is the bit rate. If the random variable $X(\gamma)$ is defined as

$$X(\gamma) = \left|\int_0^1 \exp[j\sqrt{\gamma}\psi(t)]dt\right|^2$$

(4.9)

the expression for bit error conditioned on $X(\gamma)$ is

$$P_{b_{X(\gamma)}} = E_{X(\gamma)}\left\{\frac{1}{2}\exp\left[\frac{-A^2TX(\gamma)}{4N_0}\right]\right\}$$

$$= E_{X(\gamma)}\left\{\frac{1}{2}\exp\left[-\frac{1}{2}\frac{\mathcal{E}_b}{N_0}X(\gamma)\right]\right\} \qquad (4.10)$$

From (4.9), it is clear that $0 \le X(\gamma) \le 1$. This property indicates that $X(\gamma)$ may be considered as a power fraction due to phase noise.

B.   PHASE NOISE VISUALIZATION

Significant insight into the phase noise process may be obtained from careful consideration of (4.9). It is intuitive that when the phase angle does not wander during an integration period represented by $0 \le x \le 1$, the integration is fully coherent, and therefore $X(0) = 1$. From previous discussions, the phase angle $\sqrt{\gamma}\psi(x)$ is a zero mean white Gaussian process with a variance $2\pi\beta Tx = \gamma x$ that increases throughout the integration period. The probability density for the process $Z = \sqrt{\gamma}\psi(x)$ is

$$f_z[z(x)] = \frac{1}{\sqrt{2\pi\gamma x}}\exp\left[-\frac{z^2}{2\gamma x}\right] \qquad -\infty < z < \infty \qquad (4.11)$$

A plot of the density versus phase angle and time for a generic value of γ is shown in Figure 4.2.

The phase angles of Figure 4.2 are shown in a continuous form. In actuality, since phase angle is a Modulo-2π process, the density function becomes
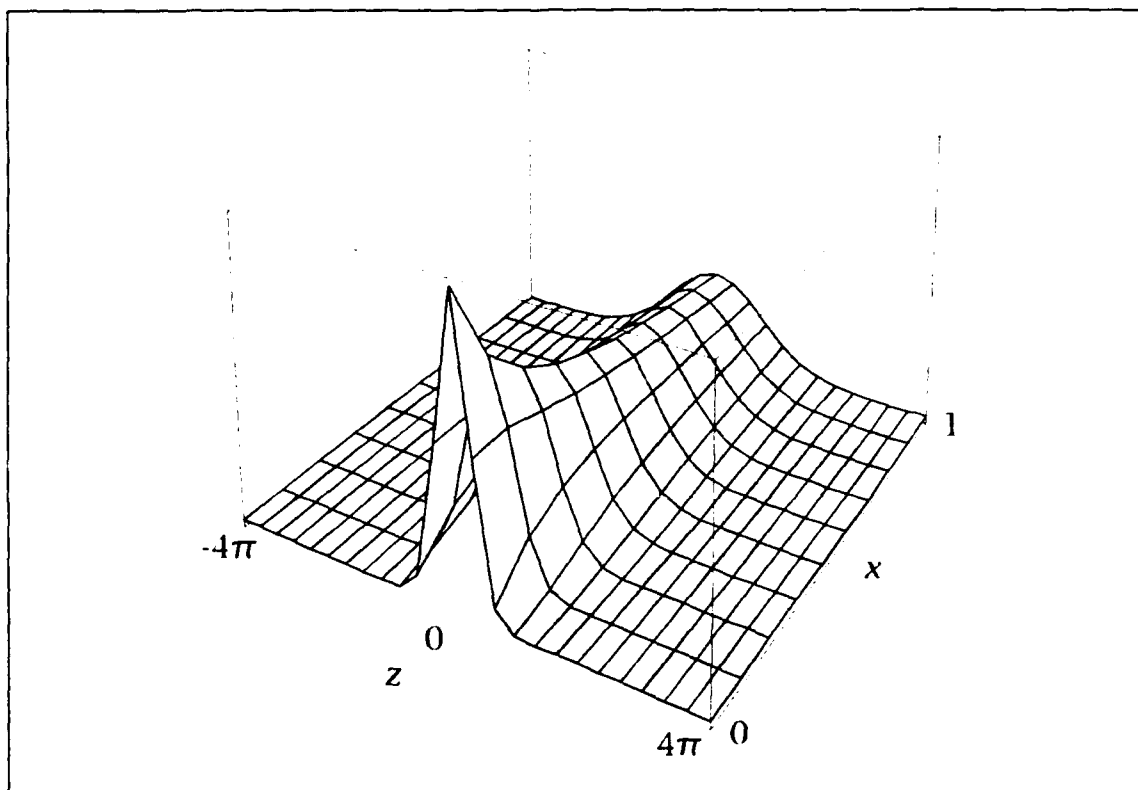
21

Figure 4.2 $f_z[z(x)]$

$$f_z[z(x)] = \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\gamma x}} \exp\left[-\frac{(z+2\pi n)^2}{2\gamma x}\right] \quad -\pi < z < \pi \qquad (4.12)$$

The density "rolls up" into the form shown in Figure 4.3. Now the impact of phase noise becomes very evident, in that as the integration period progresses, the density for the phase begins to approach a uniform density, which destroys the coherence of the integration. Figures 4.4 through 4.10 show these "rolled up" densities for a range of γ. As γ increases, the transition to a uniform density is more pronounced and rapid, leading to less time for coherent integration and lower signal power.
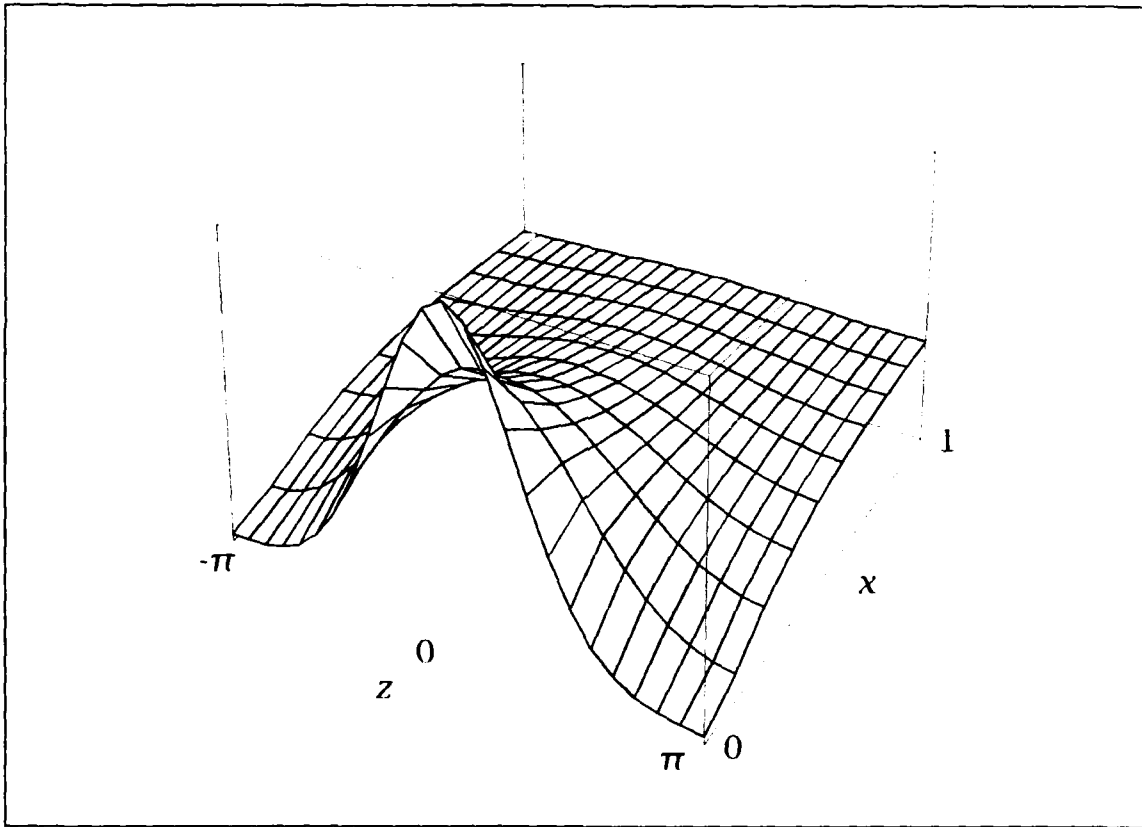
Figure 4.3 $f_Z[z(x)]$ with phase angle taken Modulo-2π
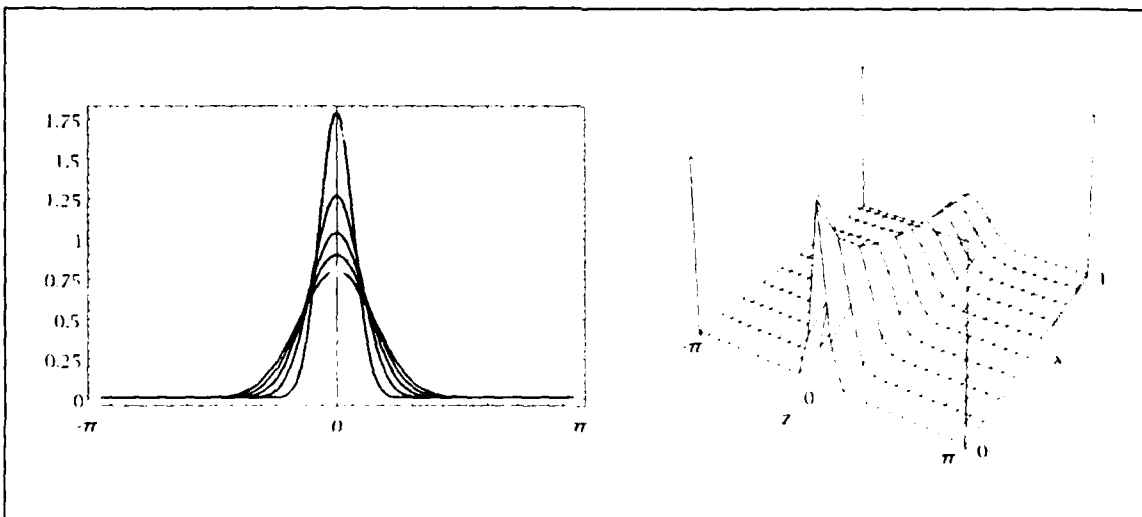

Figure 4.4 $f_z[z(x)]$ for $\gamma = 1/4$. The contours of the 2-D plot are at x

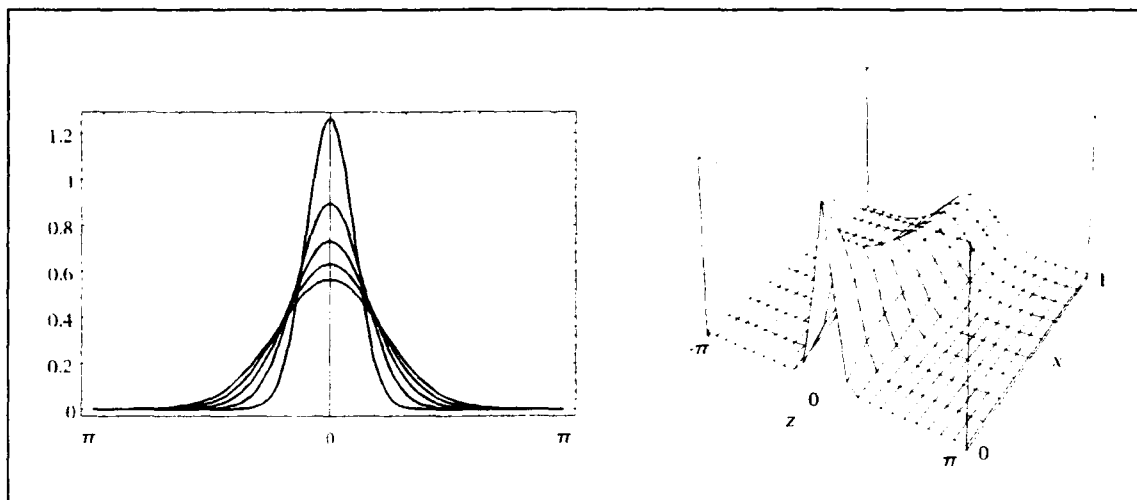intervals of $(.2,.4,.6,.8,1.0)$

Figure 4.5 $f_z[z(x)]$ for $\gamma = 1/2$.
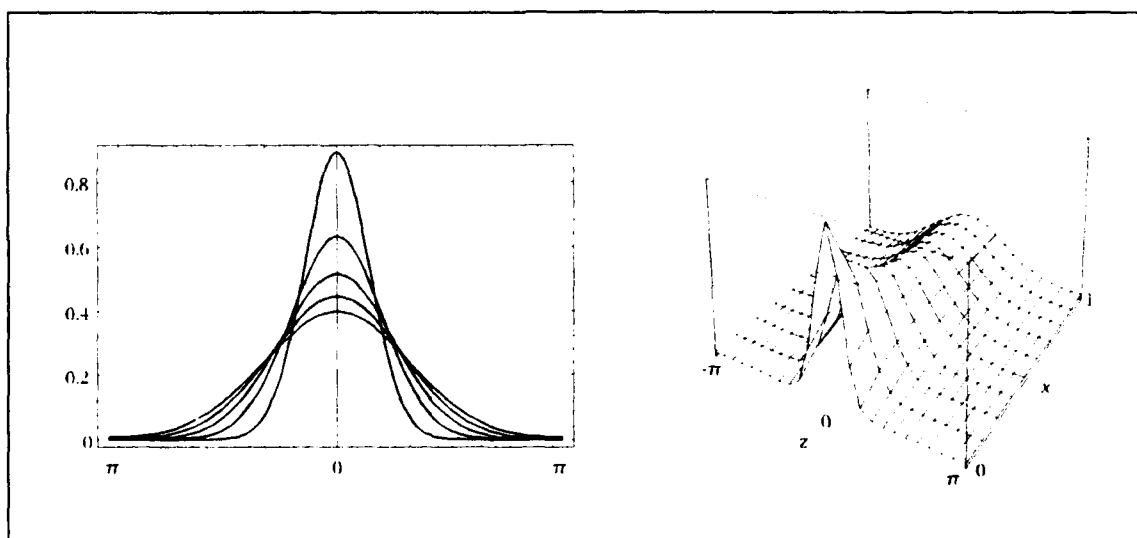
Figure 4.6 $f_z[z(x)]$ for $\gamma = 1$.
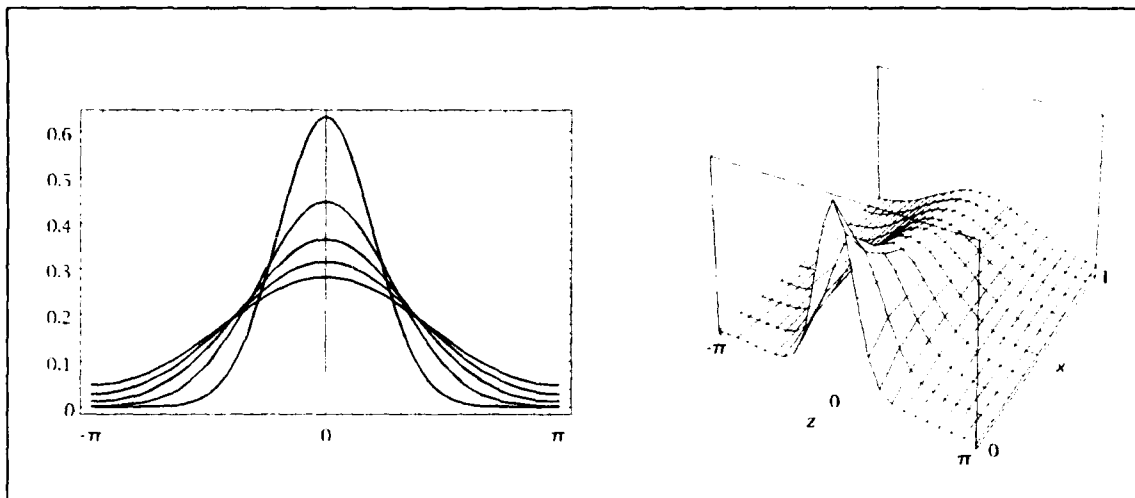
Figure 4.7 $f_z[z(x)]$ for $\gamma = 2$.
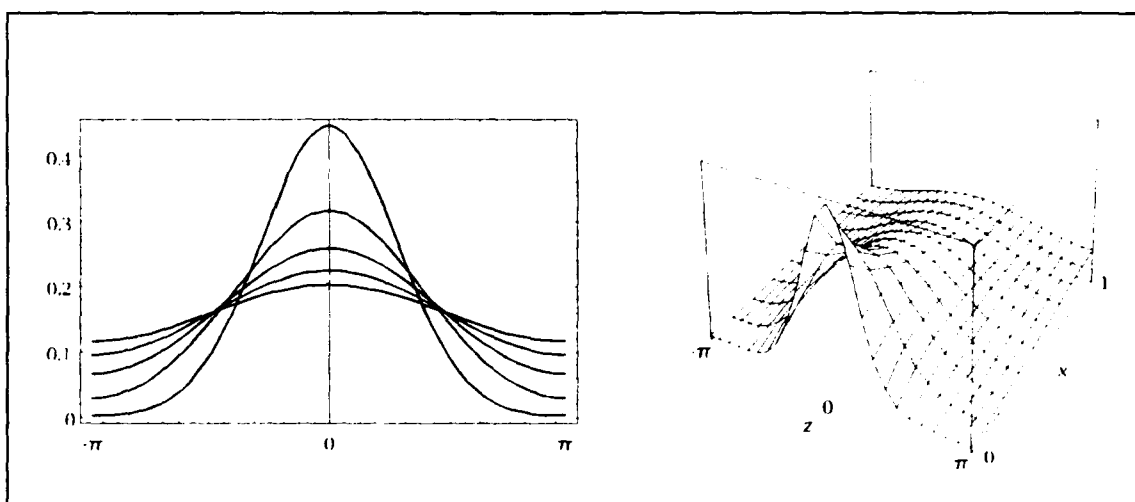
Figure 4.8 $f_z[z(x)]$ for $\gamma = 4$.
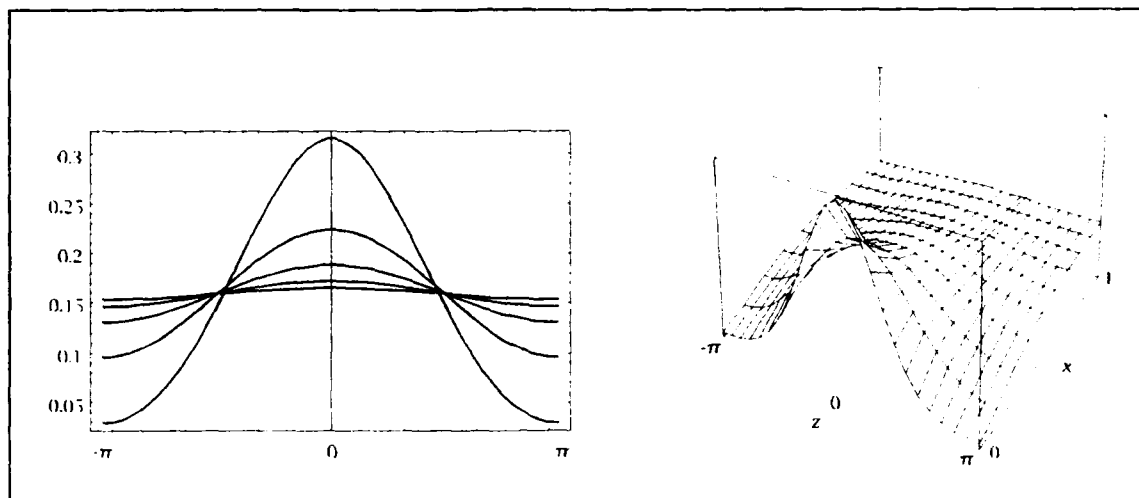
Figure 4.9 $f_z[z(x)]$ for $\gamma = 8$.
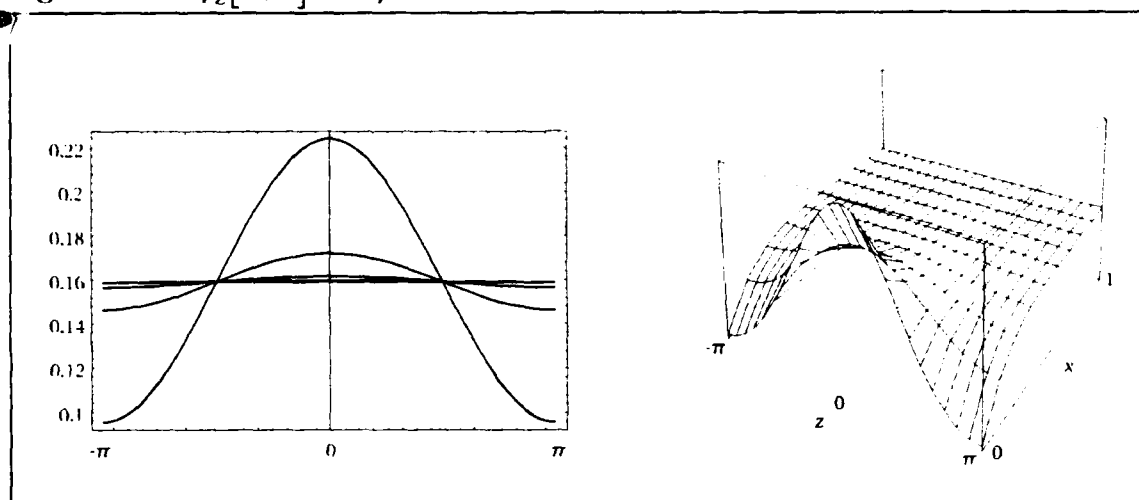
Figure 4.10 $f_z[z(x)]$ for $\gamma = 16$.

The above intuitive interpretation provides a means to establish an expected value of the random variable $X(\gamma)$. Employing (4.11), this function, call it $X_E(\gamma)$, may be defined as

$$X_E(\gamma) \equiv \left| \int_0^1 \int_{-\infty}^{\infty} \exp[jz] f_z(z) dz dx \right|^2$$

$$= \left[\frac{2}{\gamma} \left(1 - e^{-\gamma/2} \right) \right]^2$$

(4.13)

26

A plot of $X_E(\gamma)$ is shown in Figure 4.11. This figure demonstrates the degradation in signal power due to increasing phase noise.
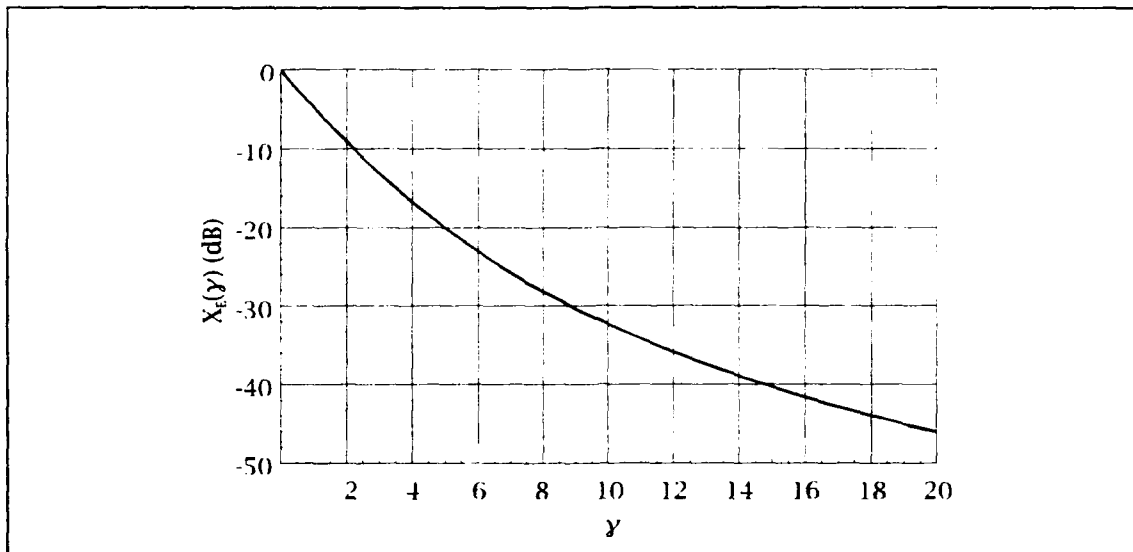
Figure 4.11 $X_E(\gamma)$

C. PHASE NOISE PDF COMPUTATION

A more traditional and complete analysis of the bit error probability for noncoherent 2FSK requires the determination of the actual probability density for $X(\gamma)$. This is a formidable problem, and to this point in time, only rough approximations to the density have been obtained.

The first approach to solving for the PDF of X was put forward by (Foschini, Greenstein, and Vannucci, 1989) and consisted of expanding the RV X in a series expansion, retaining the linear γ term

$$X_l = 1 - \gamma z^T Dz \qquad (4.14)$$

where z is an infinite dimension vector of zero mean, unit variance Gaussian random variables, and $\mathbf{D} = \text{diag}\{d_{ii}\}$, $d_{ii} = 1/(i\pi)^2$. Examination of X_L reveals that it can take on negative values, and since X represents a power fraction (with permissible values $0 \leq X(\gamma) \leq 1$), this characteristic is obviously unsatisfactory.

What is required then is an approximation to X, call it X_A, that is easy to work with analytically, has a range of values $0 \leq X_A(\gamma) \leq 1$ for all nonnegative values of γ, and closely matches the behavior of the actual variable X. Another way of stating the last requirement is that the moments of the approximation X_A match those of X. Such an approximation was formulated in (Azizoglu and Humblet, 1991), and is given by

$$X_A(\gamma) = \exp\left(-\gamma z^\mathsf{T} \mathbf{D} z\right) \qquad (4.15)$$

where z, D are as given above. The moments for the approximation are given as

$$E\{X_A^t\} = \sqrt{\frac{\sqrt{2\gamma t}}{\sinh\sqrt{2\gamma t}}} \qquad (4.16)$$

for all real $t \geq -\pi^2/2\gamma$.

Given the moments, the PDF of X_A may be determined through a Gaussian quadrature technique. The general technique is due to (Hamming, 1973), and it is believed that the application of this technique to this class of problem is new. Consider the Gaussian quadrature problem

28

$$\int_a^b K(x)f(x)dx = \sum_{k=1}^N w_k f(x_k)$$
$$K(x) \geq 0 \qquad (4.17)$$
$$a \leq x \leq b$$

If $K(x)$ represents an arbitrary probability density function of the random variable x, the moments of x are given by

$$m_t = E\{x^t\} = \int_a^b K(x)x^t dx$$
$$= \sum_{k=1}^N w_k x_k^t \qquad (4.18)$$
$$= \sqrt{\frac{\sqrt{2\gamma t}}{\sinh\sqrt{2\gamma t}}}$$

where for an N dimensional Gaussian quadrature, $t = 0, 1, \ldots 2N - 1$. Now define the following N^{th} order polynomial with roots x_k

$$\pi(x) = (x - x_1)(x - x_2)\cdots(x - x_N)$$
$$= x^N + c_{N-1}x^{N-1} + \cdots + c_0 \qquad (4.19)$$
$$c_N = 1$$

Next multiply each tth moment equation by c_t, giving

$$m_0 c_0 = \sum_{k=1}^N w_k c_0$$
$$m_1 c_1 = \sum_{k=1}^N w_k x_k c_1$$
$$m_2 c_2 = \sum_{k=1}^N w_k x_k^2 c_2$$
$$\vdots$$
$$m_{N-1}c_{N-1} = \sum_{k=1}^N w_k x_k^{N-1} c_{N-1} \qquad (4.20)$$
$$m_N = \sum_{k=1}^N w_k x_k^N$$

and add them all together

$$\left(\sum_{t=0}^{N-1} m_t c_t\right) + m_N = \sum_{t=0}^{N}\sum_{k=1}^{N} w_k x_k^t c_t$$

$$= \sum_{k=1}^{N} w_k \sum_{t=0}^{N} x_k^t c_t \qquad (4.21)$$

$$= \sum_{k=1}^{N} w_k \pi(x_k)$$

$$= 0$$

Now multiply the equation for the $(t+l)^{th}$ moment by c_t,

$$m_l c_0 = \sum_{k=1}^{N} w_k x_k^l c_0$$

$$m_{l+1} c_1 = \sum_{k=1}^{N} w_k x_k^{l+1} c_1$$

$$m_{l+2} c_2 = \sum_{k=1}^{N} w_k x_k^{l+2} c_2$$

$$\vdots \qquad\qquad (4.22)$$

$$m_{N+l-1} c_{N-1} = \sum_{k=1}^{N} w_k x_k^{N+l-1} c_{N-1}$$

$$m_{N+l} = \sum_{k=1}^{N} w_k x_k^{N+l}$$

and add them together

$$\left(\sum_{t=0}^{N-1} m_{t+l} c_t\right) + m_{N+l} = \sum_{t=0}^{N}\sum_{k=1}^{N} w_k x_k^{t+l} c_t$$

$$= \sum_{k=1}^{N} w_k x_k^l \sum_{t=0}^{N} x_k^t c_t \qquad (4.23)$$

$$= \sum_{k=1}^{N} w_k x_k^l \pi(x_k)$$

$$= 0$$

for all $l = 0, 1, \ldots N-1$. Observation of the left side of the above

equation indicates that it forms a system of N equations in N unknowns

when the moments m_i are known. In matrix notation, the system becomes

$$
\begin{bmatrix}
m_0 & m_1 & \cdots & m_{N-1} \\
m_1 & m_2 & \cdots & m_N \\
\vdots & \vdots & \ddots & \vdots \\
m_{N-1} & m_N & \cdots & m_{2N-2}
\end{bmatrix}
\begin{bmatrix}
c_0 \\
c_1 \\
\vdots \\
c_{N-1}
\end{bmatrix}
=
\begin{bmatrix}
-m_N \\
-m_{N+1} \\
\vdots \\
-m_{2N-1}
\end{bmatrix}
\tag{4.24}
$$

With the coefficients of the polynomial $\pi(x)$ in hand, the roots x_k can be numerically computed. This operation will not be considered here.

Recall the equation for the moments, repeated here

$$
\begin{aligned}
m_i &= E\{x^i\} = \int_a^b K(x)x^i dx \\
&= \sum_{k=1}^{N} w_k x_k^i
\end{aligned}
\tag{4.25}
$$

This last equation can be put into matrix form as

$$
\begin{bmatrix}
1 & 1 & \cdots & 1 \\
x_1^1 & x_2^1 & \cdots & x_N^1 \\
\vdots & \vdots & \ddots & \vdots \\
x_1^{N-1} & x_2^{N-1} & \cdots & x_N^{N-1}
\end{bmatrix}
\begin{bmatrix}
w_1 \\
w_2 \\
\vdots \\
w_N
\end{bmatrix}
=
\begin{bmatrix}
m_0 \\
m_1 \\
\vdots \\
m_{N-1}
\end{bmatrix}
\tag{4.26}
$$

Examination of the integral and summation equations for m_i indicates the correspondence to the first order of w_k and $K(x)dx$, such that

$$
\begin{aligned}
w_k &= K(x_k)\Delta x_k \\
K(x_k) &= \frac{w_k}{\Delta x_k}
\end{aligned}
\tag{4.27}
$$

In essence, this technique generates a rectangular approximation to the integral of $K(x)$. A better (smoother) method of extracting $K(x)$ from the computed w_k and x_k is to do a spline interpolation on the data

31

set, and then adjust the w_k such that the integral of the spline approximation equals 1, since any valid probability density function must integrate to 1.

Implementation of this technique is hampered by the ill conditioning of the matrix equations and root finding operation. Computation in standard 16 place arithmetic limits the N dimension of the problem to about 8. The use of a symbolic algebra system with arbitrary precision allows a larger N dimension to be used. The results of Figure 4.12 were generated using the program Mathematica with $N = 16$ and 128 digits of precision. Comparison with the results of (Azizoglu and Humblet, 1991) show equivalence. With suitable computing resources, the solution for a given N could be obtained symbolically.
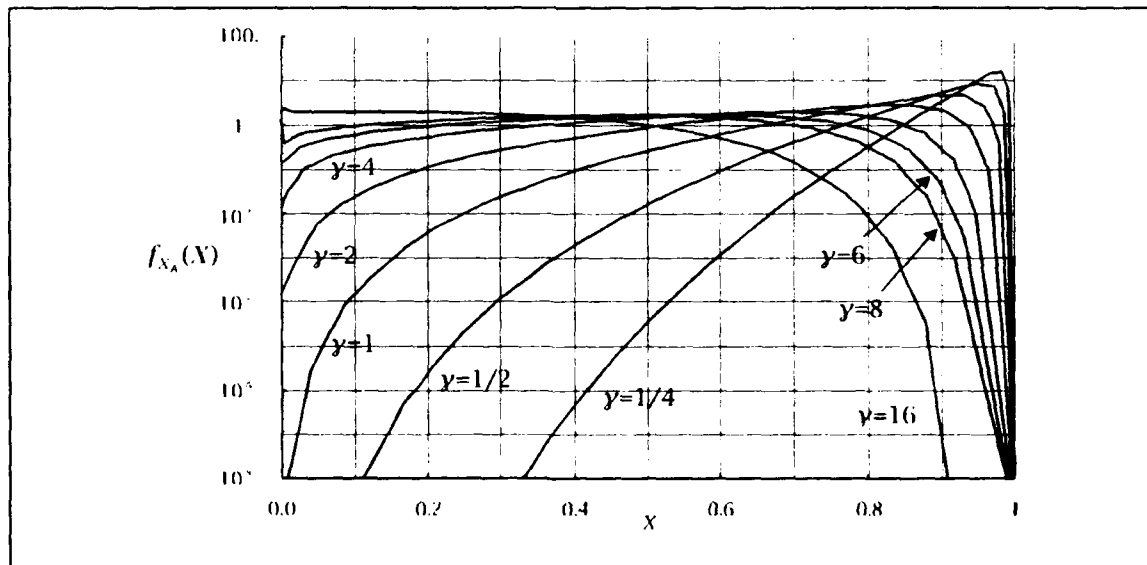

Figure 4.12 Probability Density Function of $X_A(\gamma)$

D. PERFORMANCE WITH PHASE NOISE

The density obtained for $X_A(\gamma)$ in conjunction with $P_{b_{X(\gamma)}}$ from (4.10)

yields the error performance for noncoherently demodulated 2FSK

$$P_b = \int_0^1 P_{b_{X(\gamma)}} f_{X_A}(x)dx \qquad (4.28)$$

Performance curves for several values of γ are shown in Figure 4.13.
The values for SNR represent the signal SNR, since it is assumed that one
data bit is represented by one signal bit.

The impact of increasing data rate can be explored through Figure
4.13. Starting at a point on a curve that represents a particular system,
increasing the data rate shifts the operating point to a curve at lower γ,
at a lower SNR that represents the reduced energy per bit that results
from higher data rates. For example, beginning at $\gamma = 4$, a 10^{-5} error rate
corresponds to a SNR of 32 dB. Doubling the data rate results in a drop
in SNR of 3 dB to 29 dB, on the $\gamma = 2$ curve. This point corresponds to an
improved bit error rate of 5×10^{-7}. Also consider starting at $\gamma = 1/2$, 10^{-9}
bit error rate at 18 dB. Doubling the data rate results in a worse bit error
rate of 10^{-7} on the $\gamma = 1/4$ curve at 15 dB. In general, for relatively large
γ, data rate increases lead to better error vs. SNR performance, while the
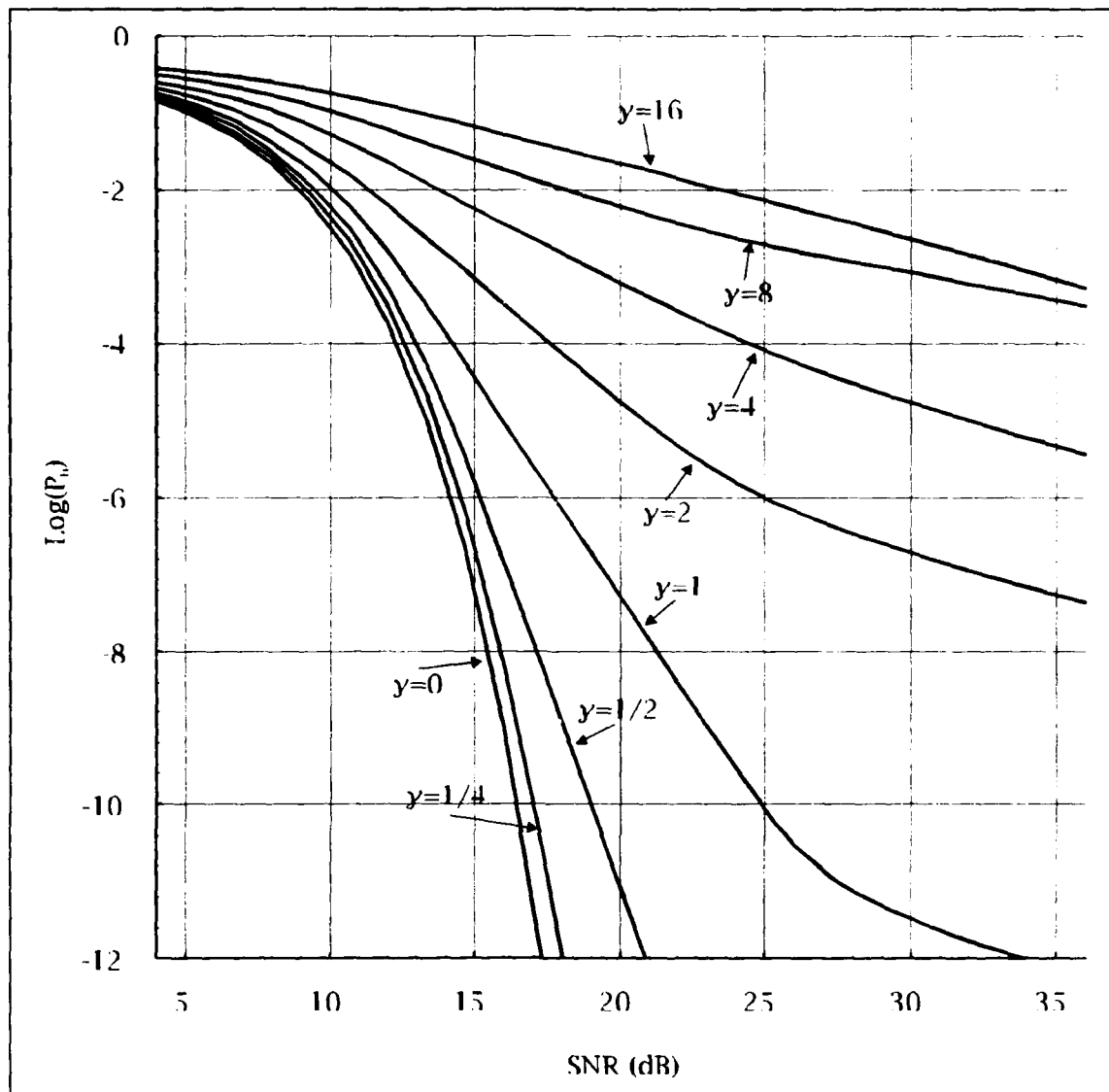converse holds at low γ.

Figure 4.13 Noncoherent 2FSK error performance

34

V. CODED PERFORMANCE

A convolutional code is generated by passing an information sequence through a linear shift register, and combining the shift register cells algebraically into several output channels. The diagram of Figure 5.1 shows a general implementation of a convolutional coder.

A. CONVOLUTIONAL CODER

Most generally, a sequence of k information bits enters the decoder at stage 1, and is subsequently shifted through the register in k bit blocks. The total number of blocks, including the input block, is L. There are a total of Lk individual cells. An actual implementation would require $(L-1)k$ memory cells, since the first stage is available directly from the coder input. The possible contents of the $(L-1)k$ coder memory cells are termed the states of the coder. There are $2^{(L-1)k}$ possible states in this coder. The input and memory cell contents (hereafter termed the coder cells) are algebraically combined via the n code generators, which are usually described as generator polynomials. More generally, the generators may be described as a sequence of Lk 1's and 0's. Where the generator is 1 indicates that the corresponding coder cell is connected to the output cell represented by this generator. Mathematically, the output of each generator is the inner product
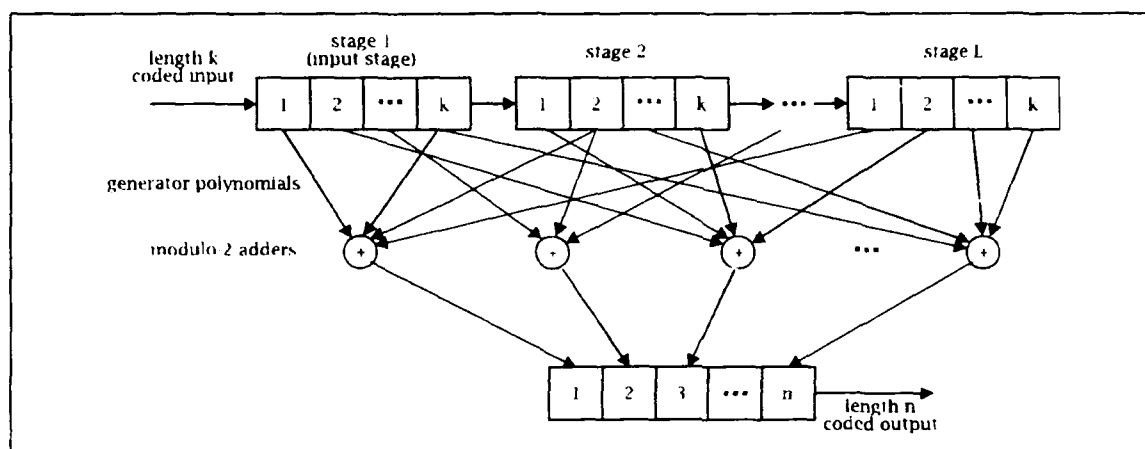
Figure 5.1　A convolutional coder

Modulo-2 of the generator and coder cells. The coder maps the k input
bits onto n output bits. The code rate of the code is

$$R_c = \frac{k}{n} \tag{5.1}$$

A convolutional coder is completely specified by the parameters L,
k, n, and the generators, which are commonly given in octal. The
mapping of the octal representation to binary is shown in Table 5.1.

TABLE 5.1

Octal	Binary
0	000
1	001
2	010
3	011
4	100
5	101
6	110
7	111

The value of L is termed the constraint length of the code. The value of Lk provides a measure of the complexity of the coder.

From observation of the coder in Figure 5.1, it is obvious that bits shifted out of the shift registers have no effect on the coder output. The constraint length therefore puts a limit on the length of the coder's impulse response (the response due to the input sequence $\{1,0,0,0,...\}$ with initial state all-zero), that being the product Lk. The impulse response $h = \{h_1, h_2, ..., h_{Lk}\}$ of the coder is analogous to that of any other linear time invariant system. For any input sequence $i = \{i_1, i_2, ..., i_z\}$, the output of the coder will be the convolution Modulo-2 of i with h.

B.   REPRESENTATION OF CONVOLUTIONAL CODES

There are three equivalent methods of describing a convolutional code: the tree diagram, the trellis diagram, and the state diagram. Each will be important to understanding the coder and describing its performance. The following example is taken from (Proakis, 1989).

Consider the $L = 3$, $k = 1$, $n = 3$ coder shown in Figure 5.2. The generator sequences are given as

$$\mathbf{g}_1 = \{1,0,0\}$$
$$\mathbf{g}_2 = \{1,0,1\} \qquad\qquad (5.2)$$
$$\mathbf{g}_3 = \{1,1,1\}$$
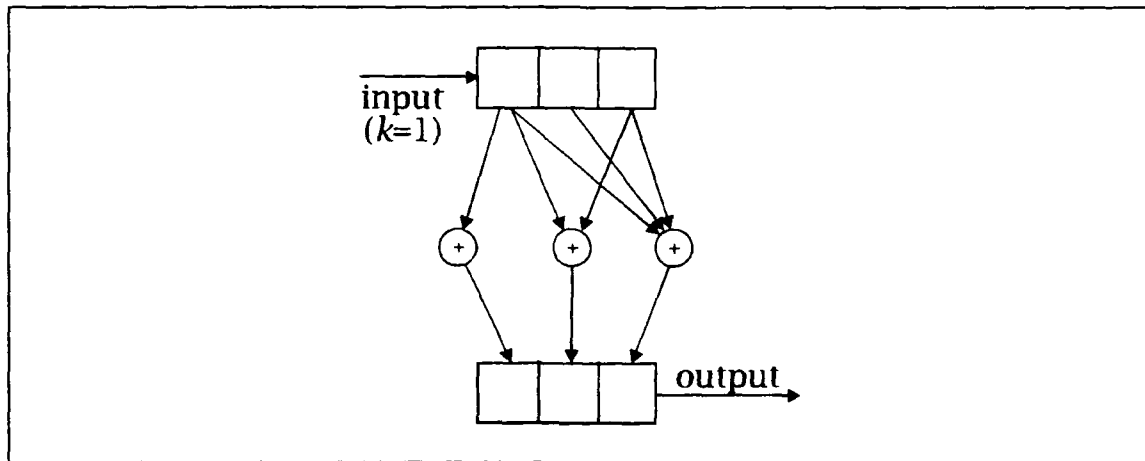
Figure 5.2　Coder for $L = 3$, $k = 1$, $n = 3$ code

In octal, these generators would be given as (4,5,7). The theory of generators frequently leads to the expression of the generator sequences as polynomials, in which case

$$g_1(x) = 1$$
$$g_2(x) = 1 + x^2 \qquad\qquad (5.3)$$
$$g_3(x) = 1 + x + x^2$$

First consider the tree diagram for this code shown in Figure 5.3. This diagram was constructed by starting at the all-zero coder state, and describes the outputs and resulting coder states for a sequence of inputs. The upper branch out of each coder state represents the transition due to an input bit 1, the lower branch to a bit 0. Starting from an all zero loading, if a bit 1 enters the coder, the state transitions to 10. The output of the coder is 111. If another 1 enters the coder, the path through the tree follows the lower branch again to state 11, and the resulting output is 110. At the input to each node in the figure, the upper number represents the coder output, and the lower number represents the coder state. The letter A represents the state 00, B

represents 01, C represents 10, and D represents 11. The upper branch out of a node is always the zero input, and the lower branch is the 1 input. This process continues for each input bit, resulting in an ever expanding diagram.

Examination of the tree diagram shows that the pattern of states and outputs begins to repeat after the third branch. Since there are a finite number of unique combinations of paths (combinations) of input, output, and coder state, the identical points on the tree diagram may be merged together. Referring to Figure 5.3, the boxed grouping of states may be merged into a single representation. When this is performed for all repeating paths, the trellis diagram results. The trellis diagram is most useful for understanding the paths of received sequences through a decoder. Figure 5.4 shows the trellis diagram for the $L = 3$, $k = 1$, $n = 3$ code. In the trellis diagram, the upper branch out of a node always corresponds to a 0 input bit, and the lower branch to a 1 input bit. The annotation next to each branch is the output sequence.

The final representation of interest is the state diagram. The state diagram is a signal flow graph showing the possible states and the transitions between them. The state diagram of Figure 5.5 shows transitions due to a 1 bit as dashed lines, and those due to a 0 input as solid lines.
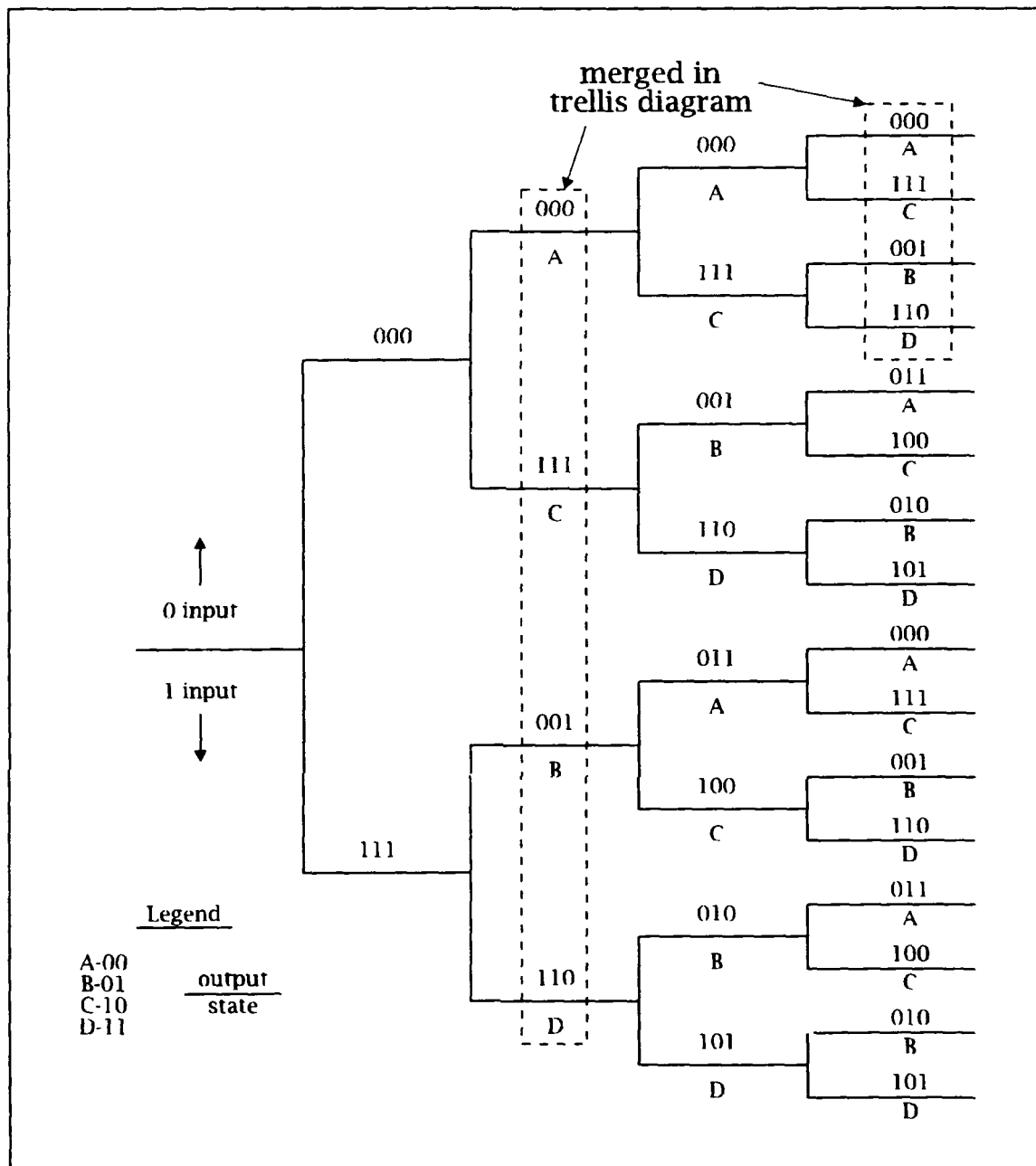
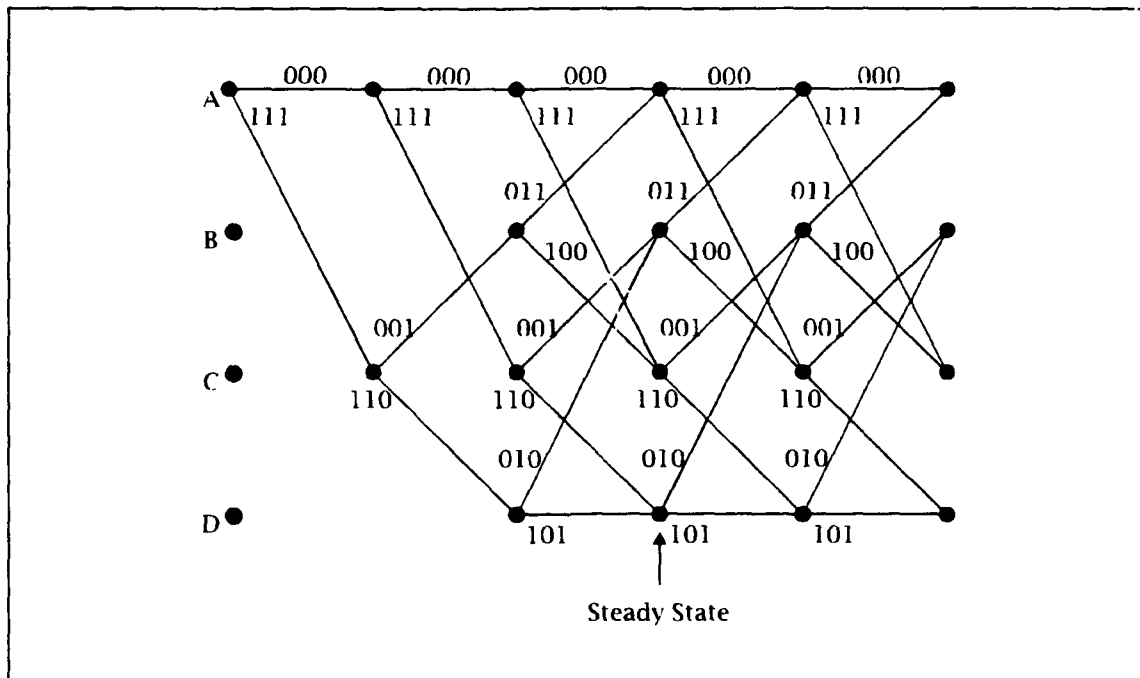Figure 5.3 Tree Diagram for $L = 3$, $k = 1$, $n = 3$ code

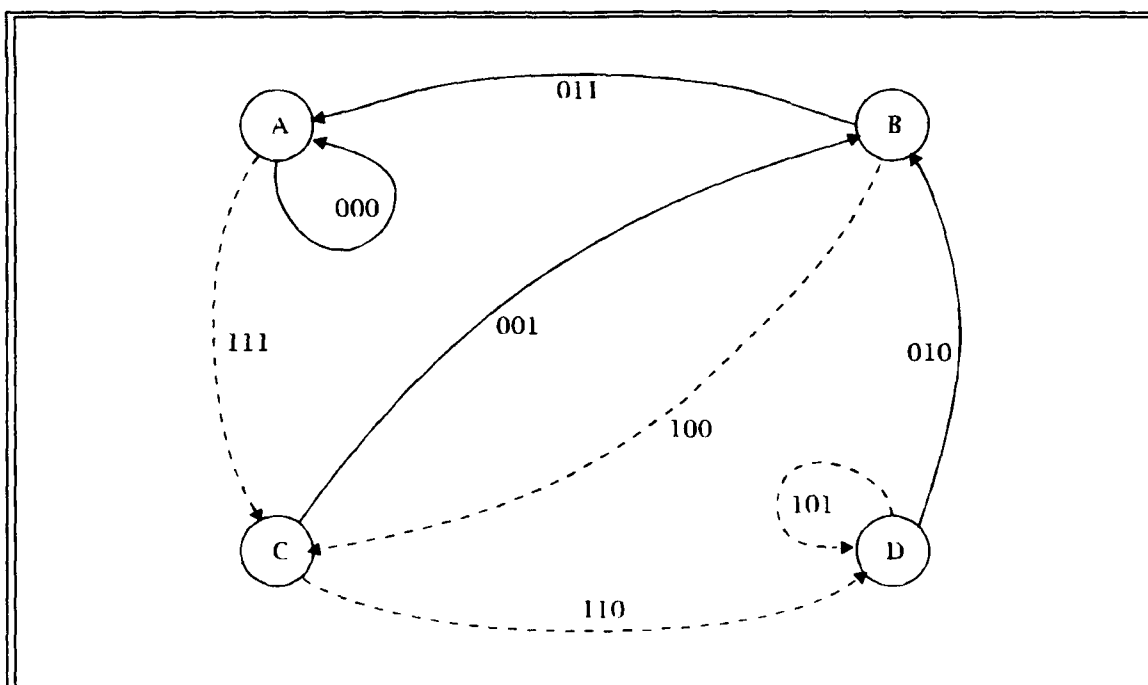Figure 5.4   Trellis Diagram for $L = 3$, $k = 1$, $n = 3$ code

Figure 5.5   State diagram for $L = 3$, $k = 1$, $n = 3$ code

41

The example depicted is for the $L = 3$, $k = 1$, $n = 3$ code. For any rate k/n, constraint length L code, there will be $2^{(L-1)k}$ possible states, each with 2^k branches entering the state, and 2^k branches leaving the state (after steady state is reached in the trellis).

The state diagram representation provides a mechanism to develop a code transfer function using standard signal flow graph techniques. The transfer function of the code provides a mechanism to determine the distance properties of the code, and to determine the error conditions that go uncorrected by the code. Since the codes being considered are linear codes (the shift register and Modulo-2 adders of the coder are linear processes), the transfer function may be constructed from the zero state to the zero state without any loss of generality. Each branch of the state diagram may be labeled with a gain function $JN^{w_i}D^{w_o}$, where w_i is the weight of the input sequence causing the transition, and w_o is the weight of the resulting output sequence. The J term serves to keep track of the number of state transitions. After labeling each branch, a set of simultaneous equations can be constructed and solved for the overall transfer function. Keeping in mind the explosion of states and inputs for large values of L and k, it can be seen that the determination of the transfer function can be difficult. The Appendix contains a more detailed discussion of the procedure, as well as a computer program for use with a symbolic algebra system to eliminate the tedium of these calculations. The state diagram with branches annotated according to the above convention is shown in Figure 5.6.
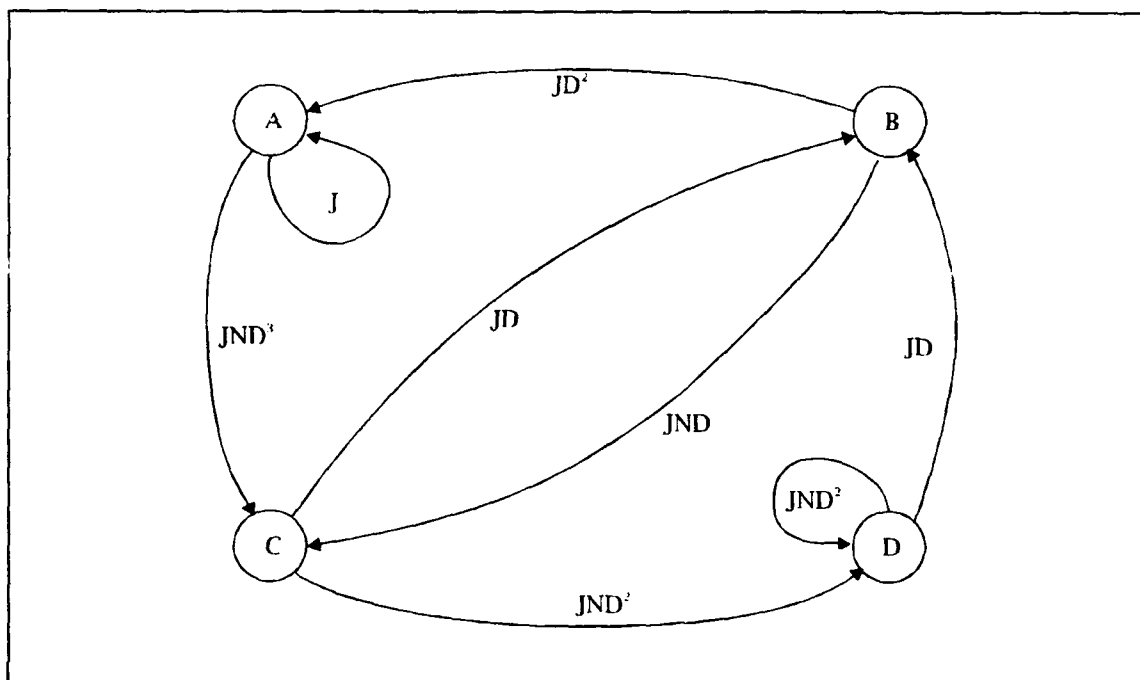
Figure 5.6   State diagram in transfer function form for $L = 3$, $k = 1$, $n = 3$ code

The system of gain functions for the example code is as follows:

$$X_B = JDX_C + JDX_D$$
$$X_C = JND^3 X_{A_{out}} + JNDX_B$$
$$X_D = JND^2 X_C + JND^2 X_D$$
$$X_{A_{in}} = JD^2 X_B$$

(5.4)

The self loop at state A is ignored in this formulation since it contributes nothing to the distance properties of the code. The system of equations is solved for $X_{A_{in}}/X_{A_{out}}$ giving

$$T(D,N,J) = \frac{J^3 ND^6}{1 - JND^2 - J^2 ND^2}$$

(5.5)

When this transfer function is expanded in Taylor series, the transfer function becomes

$$T(D,N,J) = J^3ND^6 + J^4N^2D^8 + J^5N^3D^{10} + \cdots \qquad (5.6)$$

This form reveals the distance properties of every erroneous, uncorrectable path through the decoder. The first term in (5.6) reveals that the minimum distance of the example code (the minimum exponent of D), is 6. A full explanation of the individual components of each term follows. The explanation is based on the assumption that the all zero code sequence is the transmitted sequence.

J Term: The exponent of J is the length of the information sequence with a coded representation that departs the all-zero state and subsequently remerges with it. Examination of the trellis diagram of Figure 5.4 confirms that for the example code, this path length is 3.

N Term: The exponent of N is the number of 1's in the true information sequence for the erroneous path. For the example, the erroneous path corresponds to the true information sequence 100, as can be determined from the trellis.

D Term: The exponent of D is the distance of the received coded bit sequence from the all zero sequence. For the example this is 6. The coded sequence for the example code is 111 001 011.

C. VITERBI OPTIMUM DECODING ALGORITHM

At this point, the tools required to discuss the optimum decoding of convolutional codes are in hand, and discussion will shift to this topic. The optimum decoding technique is called the Viterbi decoding algorithm. The path of quickest comprehension is by example, so the previous $L = 3$, $k = 1$, $n = 3$ code will continue to serve this function.

Consider the trellis diagram from Figure 5.4 and two of the possible paths through it

$$r_0 = \{0,0,0,0,0,0,0,0,0\}$$
$$r_1 = \{1,1,1,0,0,1,0,1,1\}$$

(5.7)

The path r_0 is the topmost path, and path r_1 is the path that remerges with the all zero path after 3 information bits. One method of decoding is to periodically take a sequence of coded bits, and determine the distance between the sequence and every possible trellis path corresponding to the sequence length. The output of the decoder would be the information sequence corresponding to the code sequence closest in distance to the received sequence. The measure of distance can be accomplished in several ways. If a hard decision is taken for each bit as to it being a 1 or 0, the distance metric is the Hamming distance. The Hamming distance is defined for binary sequences as the number of places in which the two sequences differ. The sequences r_0 and r_1 have a Hamming distance of 6. If the received bits are quantized into a number of levels without taking a firm decision as to each bit being a 1 or 0, the Euclidean distance between the quantized bits becomes the metric. This

45

approach is called soft decision decoding, and generally yields superior performance to that of hard decision decoding, at the cost of greater system complexity. This thesis will hereafter concern itself with hard decision decoding for the reasons mentioned in the introduction.

Continuing with the example, consider all paths that branch from the state at which r_0 and r_1 merge. Two sets of paths will be created, one with r_0 at the beginning, and one with r_1 at the beginning. The metrics of the extended paths will be added equally to the metrics of r_0 and r_1, preserving the difference between the metrics of r_0 and r_1. This implies that the total number of possible paths to keep track of can be split in two at the point where r_0 and r_1 merge by eliminating the path with the higher metric up to that point. Therefore, at every state where two branches merge, one entering branch can be discarded based on its metric statistics up to that point.

Now consider a very long trellis with a corresponding received sequence. The rightmost bits are the most recent inputs to the decoder. At every stage, when two candidate paths merge at a state, only one path survives. Eventually, a stage will come when only one path back to the beginning of the trellis survives. This path represents the maximum likelihood estimate of the actual transmitted data and would be the output of the decoder. Since the current coded data input is further to the right of this stage, a time delay for the decoding process is implicit. Figure 5.7 shows several steps through the procedure for the $L = 3$, $k = 1$, $n = 3$ example code. The received sequence for this example was {101 000 100 000 000 000 000}. For the case when there is a tie

46
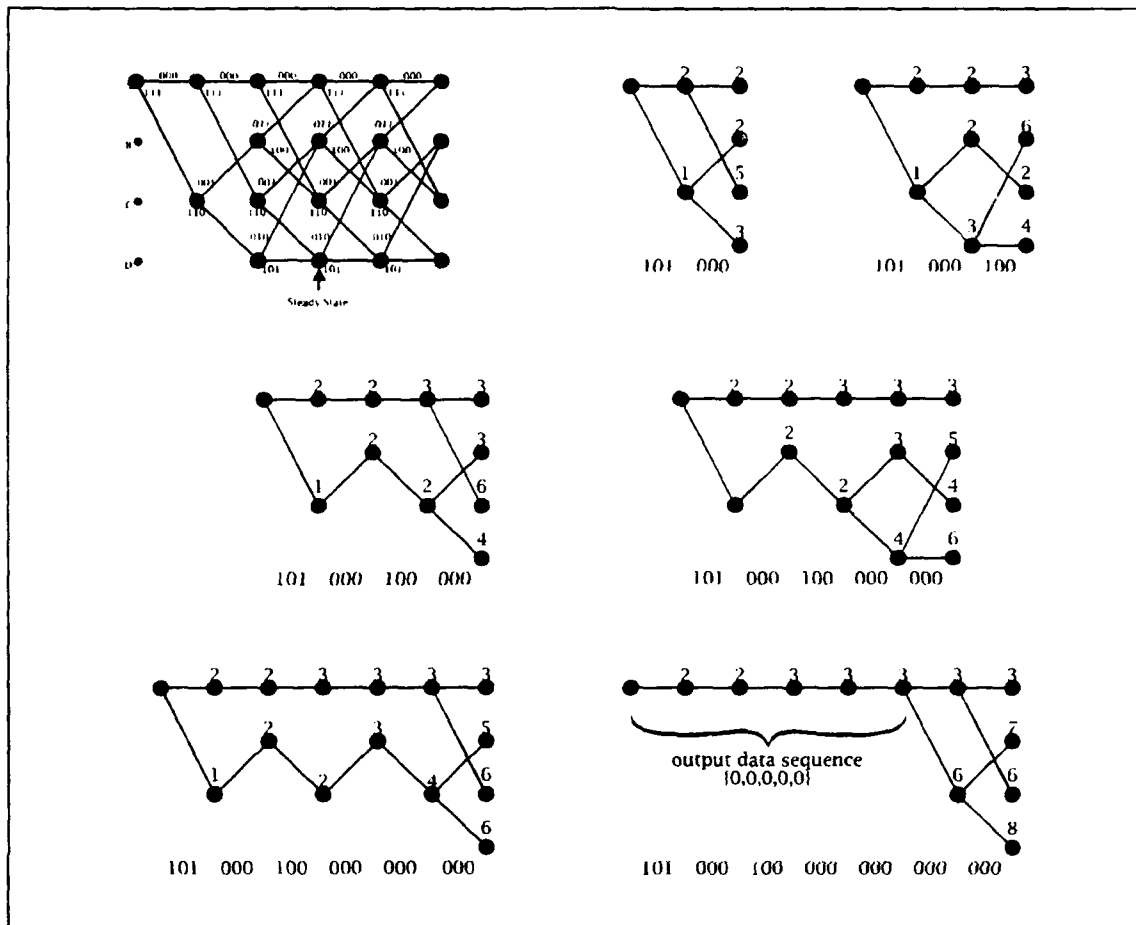
Figure 5.7 Decoding process for $L = 3$, $k = 1$, $n = 3$ code

between path metrics merging at a state, the tie is resolved in favor of the uppermost incoming branch. Each node in Figure 5.7 is labeled with the path metric up to that point.

Examination of Figure 5.7 shows how two candidate paths quickly appear. Eventually, the lower branch is eliminated, and the output sequence becomes {0,0,0,0,0}. The number of steps required to achieve a single survivor is a random variable. In practice, this number of steps is limited to a practical number, and the path with the lowest metric at the

limit is the output path. Refering to Figure 5.7, an appropriate limit would be 5 steps.

The process described above appears to be costly in that a large number of possible paths and computed metrics for each path must be kept in memory. Specifically, for a rate k/n constraint length L code, the decoder must keep track of $2^{(L-1)k}$ surviving paths with corresponding path metrics. At each stage, since there are 2^k entering branches at each state, a like number of branch metrics must be computed and added to the appropriate path metrics. These paths are then compared, and the smallest one kept. All of these mathematical operations and the memory required to store the possible paths has limited the application of this technique to codes of relatively short values of Lk, and lower data rates than is achievable through more conventional block coding techniques. The use of a soft decoding scheme increases the memory required to store the metrics, depending on the number of quantization levels, and the computation of branch metrics requires squaring and square root operations instead of simple addition operations, so even more restrictions apply to soft decoding. Nevertheless, as memory becomes cheaper, and logic speeds higher, the gap between capabilities of block and convolutional codes should decrease.

D.   CODED PERFORMANCE ANALYSIS

When hard binary decisions are taken on the incoming coded data before decoding, the communications channel is called a binary symmetric channel. In the Viterbi decoder discussed above, errors in the

48

received coded data manifest themselves in incorrect path metrics. As in the determination of the code transfer function, it may be assumed that the all zero codeword is the actually transmitted codeword without loss of generality. Recall the information that the transfer function provides about incorrect paths that remerge with the all zero path Each one of these (infinite in number) paths has a distance d associated with it. In this case, these distances are each paths' distance from the all zero code sequence. It is crucial to realize that an error in a received digit where the correct and incorrect paths are the same has equal effects on the metrics of both, and can therefore be ignored. It is this feature that permits us to consider no more than d errors even though more than this number could occur on a path. If a particular distance d is an odd number, the all zero path will be selected as the surviving path at the particular node if less than $(d+1)/2$ errors occur in the coded sequence up to that point. If $(d+1)/2$ is an even number the two paths will have equal metrics, and the probability of making the proper choice of survivor is $1/2$ when this number of errors occurs. The probability then of the incorrect path surviving is a binomial probability function

$$
P_2(d) = \begin{cases} \displaystyle\sum_{k=(d+1)/2}^{d} \binom{d}{k} p^k (1-p)^{d-k} & d \text{ odd} \\[3mm] \displaystyle\sum_{k=d/2+1}^{d} \binom{d}{k} p^k (1-p)^{d-k} + \frac{1}{2}\binom{d}{d/2} p^{d/2}(1-p)^{d/2} & d \text{ even} \end{cases}
$$

(5.8)

An upper bound can be found for $P_2(d)$ as

$$
P_2(d) < \left[2\sqrt{p(1-p)}\right]^d
$$

(5.9)

The value of p in the above equations is the channel error probability.

The above error probability is that for a single path. In general, there are a_d possible paths that merge at the particular node of the trellis. There also is an infinite number of possible merging distances after the initial d_{free}. These paths are not independent of each other, but the assumption that they are yields a pessimistic estimate for the combined error probability

$$P_e < \sum_{d=d_{free}}^{\infty} a_d P_2(d) \qquad (5.10)$$

This technique is termed the Union Bound Approach. It ignores the intersections of the different paths.

The values of a_d may be obtained from the transfer function $T(D,1,1)$. Substituting the approximation for $P_2(d)$ into the equation for path error probability yields

$$P_e < T(D,1,1)\Big|_{D=2\sqrt{p(1-p)}} \qquad (5.11)$$

The statistic of greatest importance is the information bit error rate, which may be obtained as follows. When an incorrect path is chosen, the information bits in that path will differ from the transmitted information. Since the all zero information path was the one transmitted, the number of errors will be the number of 1's in the incorrectly chosen path. The information bit error probability for a single path is the product of the path error and the number of 1's in the erroneous information sequence. Consider the transfer function in the following form with the a_{d_i} representing the number of paths of weight d with information weight w_i

$$T(D,N,1) = \sum_{d=d_{free}}^{\infty} D^d \sum_i a_{d_i} N^{w_{i}}$$

$$= \sum_{d=d_{free}}^{\infty} a_d D^d N^{f(d)} \tag{5.12}$$

The mean number of 1's in the information sequence corresponding to a path is given by $f(d)$. The information bit error probability for all paths of a given weight may be over bounded by the mean

$$P_b < f(d) P_e$$

$$< \sum_{d=d_{free}}^{\infty} a_d f(d) P_2(d) \tag{5.13}$$

Recognizing that

$$\frac{\partial T(D,N)}{\partial N} = \sum_{d=d_{free}}^{\infty} a_d D^d f(d) N^{f(d)-1} \tag{5.14}$$

and employing the bound on $P_2(d)$ of (5.9), the bit error probability becomes

$$P_b < \frac{\partial T(D,N)}{\partial N}\bigg|_{N=1,D=2\sqrt{p(1-p)}} \tag{5.15}$$

This expression for P_b is the expression that will be used to compute the performance of the laser phase noise contaminated binary FSK system described earlier.

E. CODES STUDIED

Four codes will be considered. The first code is the $R = 1/2$, $L = 3$ code with generators (in octal) of (5,7) (Odenwalder, 1970). The transfer functions for this code are

$$T(D,N,J) = \frac{D^5 NJ^3}{1 - DNJ^2 - DNJ} \qquad (5.16)$$

$$\left. \frac{\partial T(D,N)}{\partial N} \right|_{N=1} = \frac{D^5}{(2D-1)^2} \qquad (5.17)$$

$$= D^5 + 4D^6 + 12D^7 + 32D^8 + 80D^9 + 192D^{10} + \cdots$$

The next code is the $R = 1/2$, $L = 4$ code with generators (15,17)

(Odenwalder, 1970). The transfer functions are

$$T(D,N,J) = \frac{D^6 J^4 N \left(D + JN - D^2 JN \right)}{1 - DJN - DJ^2 N - D^3 J^3 N + D^2 J^3 N^2 - \cdots} \qquad (5.18)$$
$$D^4 J^3 N^2 - D^2 J^4 N^2 + D^4 J^4 N^2$$

$$\left. \frac{\partial T(D,N)}{\partial N} \right|_{N=1} = \frac{2D^6 - D^7 - 2D^8 + D^9 + D^{11}}{\left(D^3 + 2D - 1 \right)^2} \qquad (5.19)$$

$$= 2D^6 + 7D^7 + 18D^8 + 49D^9 + 130D^{10} + 333D^{11} + \cdots$$

The third code is the $R = 1/4$, $L = 3$ code with generators (5,7,7,7)

(Larsen, 1973) and transfer functions

$$T(D,J,N) = \frac{J^3 N \left(D^{11} + D^{10} JN - D^{14} JN \right)}{1 - D^3 JN - D^3 J^2 N - D^2 J^3 N^2 + D^6 J^3 N^2} \qquad (5.20)$$

$$\left. \frac{\partial T(D,N)}{\partial N} \right|_{N=1} = \frac{2D^{10} + D^{11} - D^{13} - 2D^{14} + D^{17}}{\left(D^6 - 2D^3 - D^2 + 1 \right)^2} \qquad (5.21)$$

$$= 2D^{10} + D^{11} + 4D^{12} + 9D^{13} + 8D^{14} + 25D^{15}$$

The final code is the $R = 1/4$, $L = 4$ code with generators (13,15,15,17)

(Larsen, 1973) and transfer functions

$$T(D,J,N) = \frac{D^{13} J^4 N \left(1 + DJN - D^3 JN + J^2 N^2 - D^4 J^2 N^2 \right)}{1 - D^3 JN - D^3 J^2 N - D^5 J^3 N + D^6 J^3 N^2 - D^8 J^3 N^2 - D^6 J^4 N^2 + \cdots}$$
$$D^8 J^4 N^2 - D^7 J^4 N^3 + D^{11} J^4 N^3 - D^5 J^5 N^3 + D^9 J^5 N^3$$

$$(5.22)$$

$$\frac{\partial T(D,N)}{\partial N}\bigg|_{N=1} = \frac{4D^{13} + 2D^{14} - 6D^{16} - 5D^{17} + 2D^{19} + 6D^{20} + \cdots}{\left(D^{11} + D^9 - D^7 - 2D^5 - 2D^3 + 1\right)^2}$$

$$= 4D^{13} + 2D^{14} + 10D^{16} + 3D^{17} + 16D^{18} + 34D^{19}$$

$$(5.23)$$

These codes are of relatively short constraint length; the current standard in commercial production for microwave radio use is a constraint length of 7. The intended application for these codes is at a high data rate, so it seems prudent to choose codes that offer lower complexity of implementation.

F.   CODED PERFORMANCE

The computation of information bit error proceeds as outlined above, realizing that the value of p comes from the development for uncoded phase noisy binary FSK. When coding is employed, since the signal bit duration T decreases by the factor R, the signal to noise ratio also decreases by R. Also, and most importantly, the value for γ decreases by R. The error performance curves based on uncoded γ values of 1, 4, and 16 are shown in Figures 5.8 through 5.10. To facilitate direct comparisons with the uncoded system, the SNR values are with respect to the uncoded system. The actual signal SNR is reduced by the code rate.

The traditional additive white Gaussian noise results for coding show that code constraint length is the dominant factor in improved performance since larger L produces a greater d_{free}. The results shown in

53

Figures 5.8 through 5.10 show that code rate is the dominant performance factor in the presence of phase noise. This code rate effect comes about from the replacement of a single information bit, with its mostly non coherently integrating latter portion, by several code bits, which integrate more coherently.

Figures 5.8 through 5.10 show an extremely large coding gain. Performance of this quality enables the realization of practical heterodyne systems, operating at 50 - 100 Mbps per wavelength channel employing the new high speed Viterbi (hard decision) decoders. With current laser diode sources exhibiting 10 - 100 MHz linewidths, a 200 Mbps signal rate results in $1 < \gamma < 4$. It should be stated that the bounds employed in the coded bit error are somewhat loose bounds, resulting in a pessimistic coded error rate.
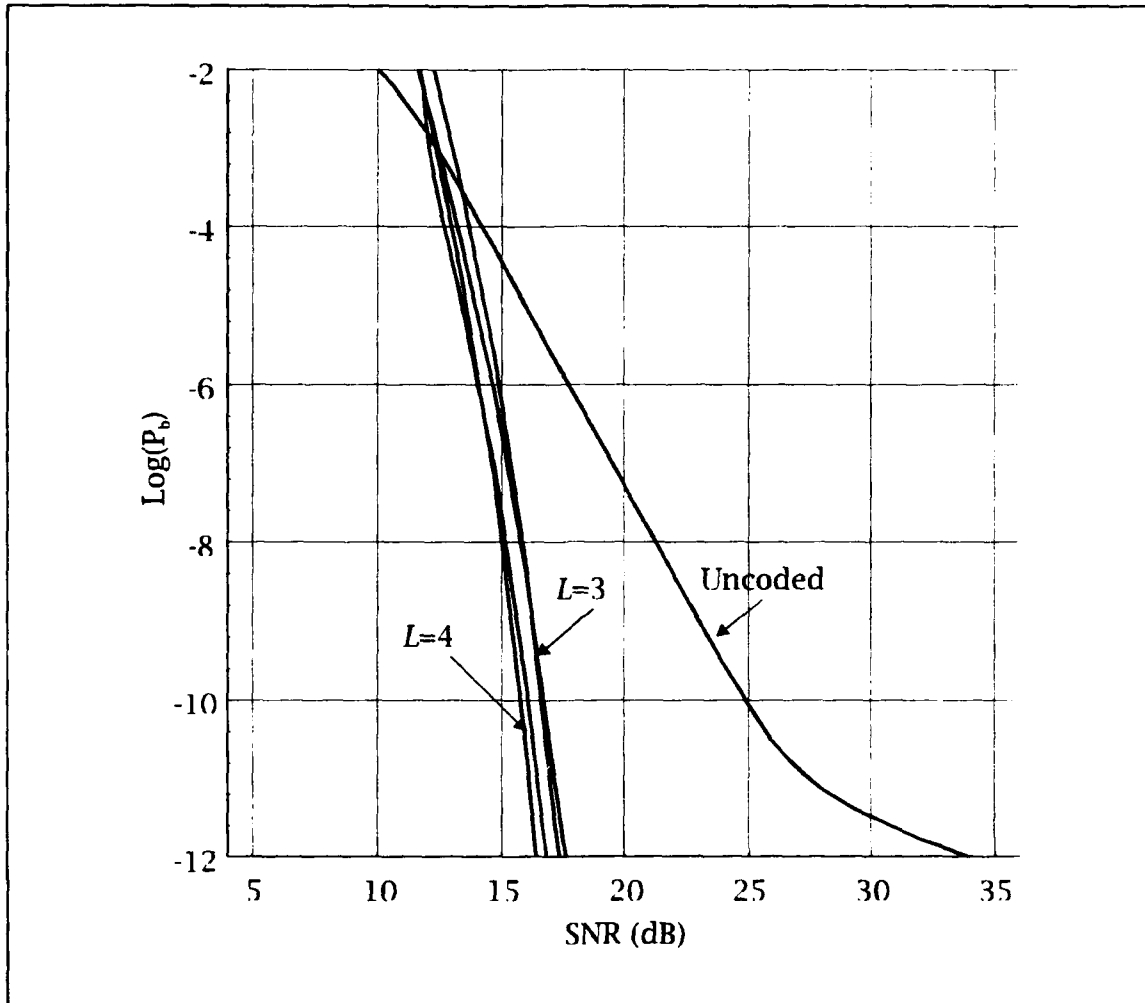
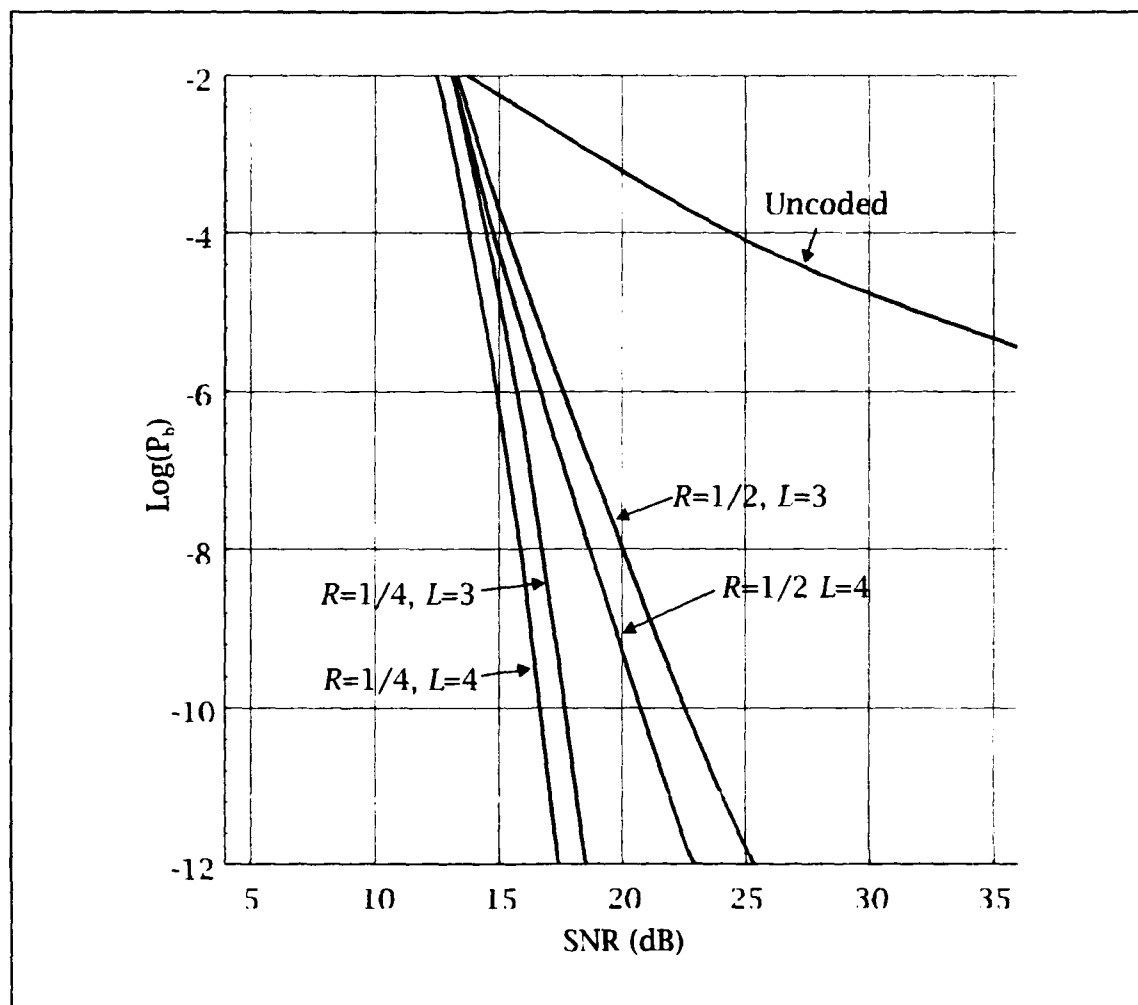Figure 5.8   Coded Performance $\gamma = 1$
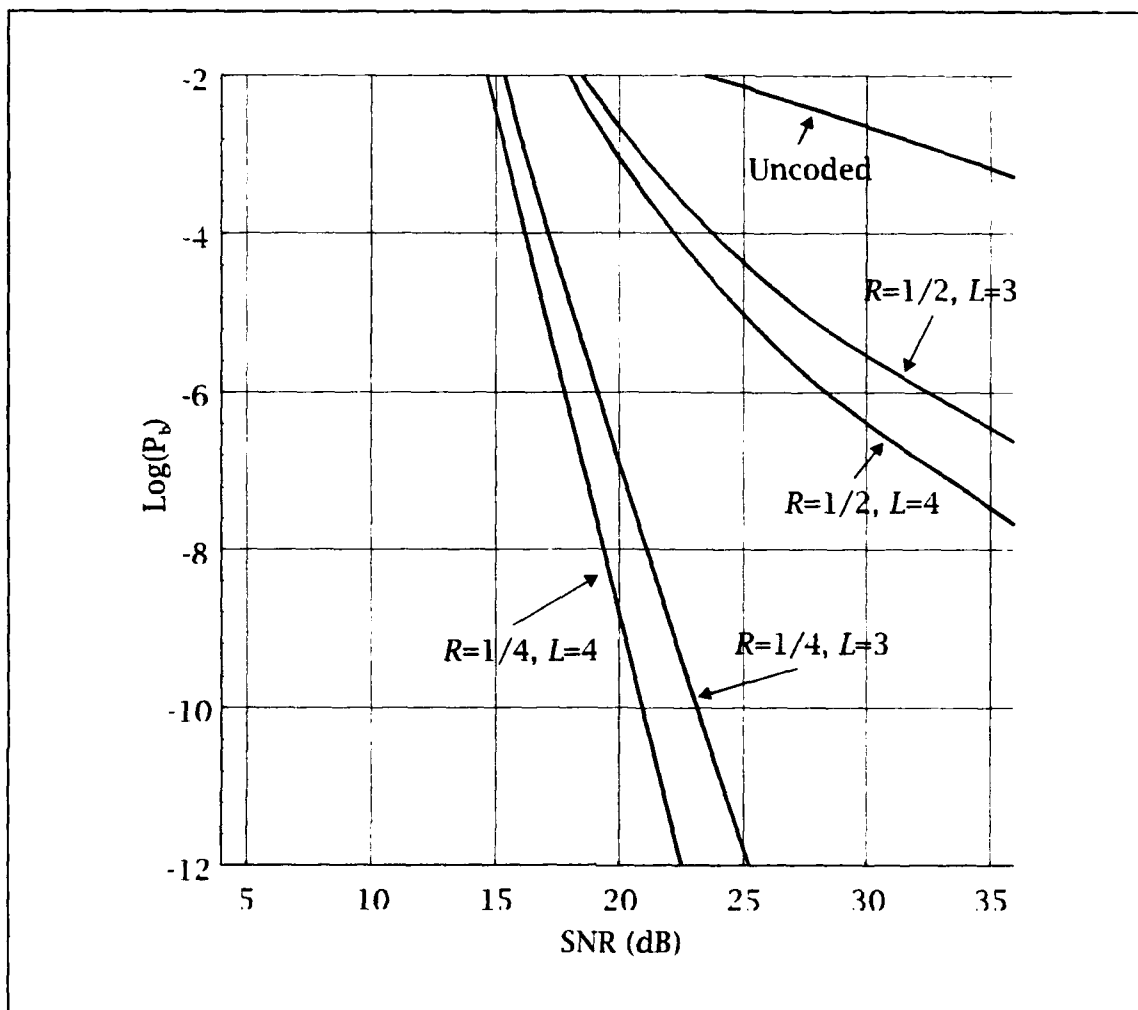
Figure 5.9   Coded Performance $\gamma = 4$

Figure 5.10 Coded Performance $\gamma = 16$

VI. CONCLUSION

The comments made in the introduction about the tradeoffs that arise from the use of coding can now be considered in more detail. It seems clear that in the absence of any constraints on signal rate and decoder performance, coding offers superior performance. Given constraints, the choice to use coding may become less clear-cut.

If higher data rates are not required, and the source laser and decoders support the higher signal rate coding requires, then coding offers significant performance gains as compared to a system that increases the signal rate without using the increased data rate available. For the uncoded system, this is equivalent to having unused data capacity on the link. Table 6.1 summarizes the required signal SNR for given bit error rates and γ values employing uncoded, coded, and uncoded signal-rate-increased systems. The γ values are referenced to the uncoded system.

A. HIGH DATA RATES VERSUS LOW DATA RATES

Experimental coherent systems have been run at modulation rates well over 1 Gbps. The fastest decoders are currently limited to 200 Mbps. It will be assumed that the combined laser linewidth of source and local lasers is 100 MHz or less. A data rate of 200 Mbps corresponds to a γ of about 4. The performance of an uncoded system is therefore

TABLE 6.1. SIGNAL SNR REQUIRED

System	$\gamma = 1$			$\gamma = 4$			$\gamma = 16$		
	10^{-5}	10^{-7}	10^{-9}	10^{-5}	10^{-7}	10^{-9}	10^{-5}	10^{-7}	10^{-9}
Uncoded	16	19	23	32	-	-	-	-	-
$L = 3$, $R = 1/2$ Code	11	12	14	14	16	19	24	34	-
$L = 3$, $R = 1/4$ Code	8	9	11	9	10	11	12	14	16
2x Signal Rate	14	16	18	21	32	-	-	-	-
4x Signal Rate	13	15	17	16	19	23	32	-	-

quite bad. In this case, a signal rate increase is the only option, unless one considers multiplexing coded systems. The total power required in the multiplexed systems to achieve a given bit error rate is less than that required in the increased signal rate system, but at significantly greater complexity. In general, the choices to be made for high data rate systems are not easy ones.

For low data rate systems, the choice does appear to be easy. So long as a decoder for the data rate desired is available, coding offers a solution to the phase noise problem.

The critical components in all this are the source and local lasers. It is a very high priority within the communication industry to develop the perfect laser for coherent communications. Work is proceeding slowly toward this goal. Semiconductor lasers offering linewidths less than 10 MHz, and modulation rates higher than 500 MHz are currently advertised.

Given a perfect laser, coding will still find application to combat other sources of noise in communication systems. The original applications of coding were to combat additive white Gaussian noise in communication channels. As long as economical decoders operating at high enough data rates are available, coding will find application.

B.   COMPARISON WITH DIRECT DETECTION

Since it has been established that a coded coherent system is at least theoretically able to offer excellent performance, it is worthwhile to consider how well such a system compares to the traditional direct detection system. The first step is to consider the error rate versus SNR of the direct detection system (Keiser, 1991).

$$P_b = \frac{1}{2}\text{erfc}\left(\frac{1}{2}\sqrt{\frac{\mathcal{E}_b}{2N_0}}\right) \tag{6.1}$$

This performance is plotted in Figure 6.1.

The direct detection error performance is roughly equivalent to uncoded coherent 2FSK with $\gamma < 1$. When comparing with the coded 2FSK systems, it is seen that the $R = 1/4$ codes perform as well or better for all $\gamma \le 16$, and the $R = 1/2$ codes perform as well or better for all $\gamma \le 4$. It appears as though the 10+ dB receiver sensitivity advantage of coherent systems may now be realized in the face of phase noise. Coding may be employed on the direct detection system, but the tremendous gains of the coherent system will not be realized, since the main factor in that effect is the effective reduction in γ brought about by

60

system are likely to be equivalent to the well-known AWGN channel results (about 6 dB), so the phase noise contaminated coded system still offers superior performance.
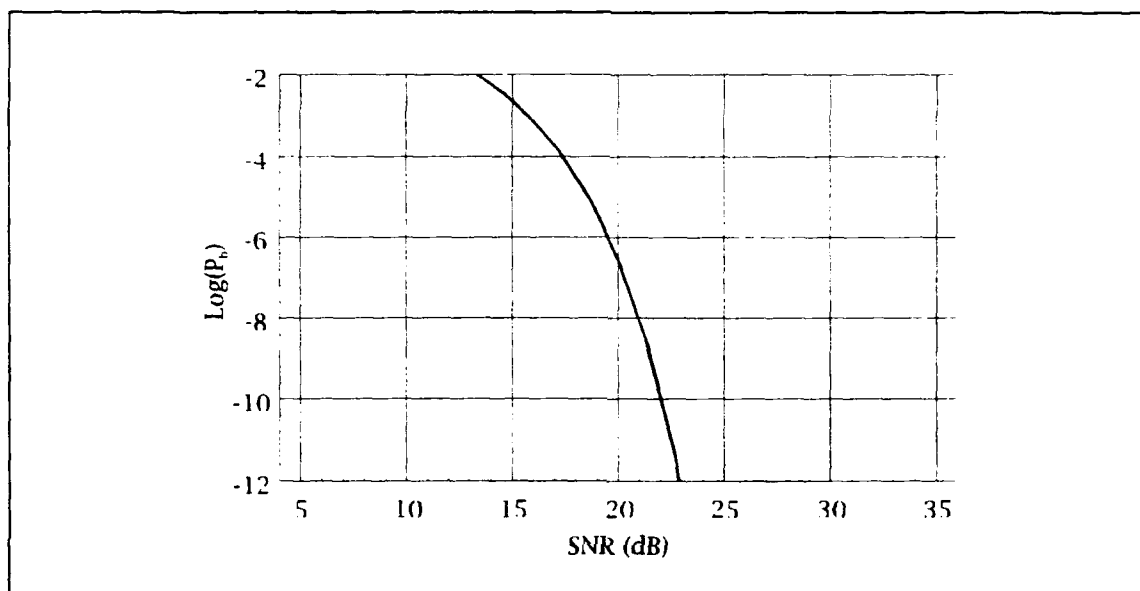
Figure 6.1   Direct Detection P_b

C.   AN APPLICATION AND FUTURE RESEARCH

This thesis has shown that good performance is attainable at low data rates with a coded coherent binary FSK system. An attempt has been made to point out at the pertinent places some of the implementation issues facing the employment of such systems. A hypothetical application of such a system will now be discussed.

There are military applications of fiberoptic links that have been conceived not so much for the high data rates available, but for the light

weight, high tensile strength, and RFI immunity aspects of fiber. Consider a link between a low value (or expendable) asset and a receiver station. Consider a medium range system (50 Km) that requires a relatively low data rate (less than 50 Mbps). There may also be a link from the receiver station to the low value asset (LVA), but it is not of interest (it would make use of the same fiberoptic cable). What is required in the LVA is a simple, inexpensive, low power consumption optical transmission system. The cost and complexity of the receiving station are of secondary importance.

Before deciding whether a coherent or direct detection system should be used, consider the following. The range requirement and desire for the lowest power system mandate the use of single mode fiber. Chromatic dispersion effects also need to be minimized, dictating the use of a laser diode source as opposed to an LED (Keiser, 1991, pp. 326-327). As a result, the two major cost and simplicity factors favoring a direct detection system are no longer part of the equation.

The choice between coherent and direct detection systems now becomes whether the complexity and cost of the coherent system justify the better performance. Earlier it was stated that the frequencies of the source and local oscillator lasers fluctuate with temperature and small drive current fluctuations. If the controllers for these parameters can be removed from the LVA, then most of the complexity problem within the LVA disappears. To remove these components may require a special, adaptive local oscillator laser and demodulator section within the

62

receiving station, and the effects on performance of such a scheme requires further investigation.

Polarization mismatch between the source and local lasers is also a factor. In general, any twisting or bending of a fiber will alter the polarization of the signal. Receivers that are insensitive to polarization also require further investigation.

Current single mode fibers offer attenuation's as low as 0.3 dB/Km. This low attenuation, coupled with the sensitivity advantages of coherent systems, offers tremendous increases in the operational range of systems that employ this concept of operation. Given that the hurdles of phase noise, frequency stability, and polarization sensitivity can be overcome, entirely new classes of weapon and sensor systems may become reality.

APPENDIX

The following Mathematica codes (Wolfram, 1992) implement the formulations for:

The Gaussian quadrature method for the determination of the phase noise pdf

Code transfer functions.

A.   GAUSSIAN QUADRATURE METHOD FOR PDF DETERMINATION FROM MOMENTS.

The first section defines the moments and structure of the matrices involved.

```
u[g_,i_]:=Sqrt[ Sqrt[2 g i]/Sinh[Sqrt[2 g i]]]
u[g_,0]:=1
MA[g_,n_]:=Table[u[g,j+k],{j,0,n-1},{k,0,n-1}];
MB[g_,n_]:=Table[-u[g,k],{k,n,2n-1}];
X[n_]:=Table[x^i,{i,0,n}];
MD[xk_List]:=Table[xk^i,{i,0,Length[xk]-1}];
ME[g_,n_]:=Table[u[g,k],{k,0,n-1}];
```

The next section computes the x_k and w_k

```
prec=128;              ... for 128 digits of precision

n=16;                  ... for N = 16

g=1;                   ... for γ = 1

A=N[MA[g,n],prec];... Remove the N[•,prec] statements for symbolic
                              answers
B=N[MB[g,n],prec];

CC=LinearSolve[A,B];

AppendTo[CC,1];

poly=CC . X[n];

XK=x /. NSolve[poly==0,x,prec];... use Solve for symbolic
                              answers
DD=N[MD[XK],prec];

EE=N[ME[g,n],prec];

WK=LinearSolve[DD,EE];

F=Transpose[{XK,WK}];
```

The data generated by the code above is next exponentially
interpolated (giving straight lines on a log plot), and an
`InterpolatingFunction` object is obtained. This object is numerically
integrated to find its norm, and finally divided by its norm, so that it
integrates to 1.

```
Fl=F /. {a_,b_}->{a,Log[b]};

Fi=Interpolation[Fl[g],InterpolationOrder->1];

norm=NIntegrate[Exp[Fi[x]],{x,0,1}];
```

```
F[g,x_]:=Exp[Fi[x]]/norm;      ... The final PDF
Plot[F[g,x],{x,0,1}];               ... Plot PDF
```

B. DETERMINATION OF CODE TRANSFER FUNCTIONS

Algorithm

1. Separate the all zero state into two states, x_i and x_f, such that transitions *to* the all zero state are attached to the x_f state, and all transitions *from* the all zero state are attached to x_i.

2. For every possible state (there will be $2^{(L-1)k}$ of them), determine the coder output sequence that occurs for every possible input sequence of length k (there will be 2^k of them). Recall that the coder output sequence is the dot product of the generator sequence with the length Lk input and shift register sequence (mod-2). Also determine the resulting state for the given input sequence.

3. Connect the original state to the resulting state with an arrow, and annotate the connection by the gain functional $LN^{w_i}D^{w_o}$ where w_i is the weight of the input sequence, and w_o is the weight of the corresponding output sequence.

4. Keep track of these transitions and path gains, and form a system of equations that relate each state as a sum of other states with corresponding gains. For given L and k, there will be $2^{(L-1)k}$ equations in $2^{(L-1)k} + 1$ unknowns, since the all zero state was split into the states x_i and x_f. Each equation should be in the form

$X_j = g_{j1}X_{j1} + g_{j2}X_{j2} + \cdots + g_{j2^k}X_{j2^k}$ where the subscripted states and gains refer to other states. In other words, each equation has 2^k components.

5. Solve the system of equations by eliminating all states except for x_i and x_f, and solve for the ratio x_f/x_i.

It is obvious that the task of finding the transfer function for typical codes is impossible without the aid of a computer. Even with a symbolic math system, the problem is daunting for large values of Lk. The above algorithm has been implemented as a Mathematica program, and used to determine the code transfer functions employed in this research. The code is listed below.

The code is given as a Mathematica Function, the assumed inputs are the values of L, k, and the generators in the form of a matrix with the rows representing the generators in binary form.

```
BinaryConvCodeTransFunc[L_Integer,K_Integer,

                          generators_?MatrixQ]:=

Module[{},

  (* L is constraint length *)

  (* K is input block length *)

  n=Length[generators];

  (* n is number of generators *)

  numstates=2^(K*L-K);

  ls=numstates+1;

  M=2^K;
```

```
xsol=Table[{0},{ls}];

allstates=Table[

                Take[Join[Table[0,{K*(L-1)}],

                        IntegerDigits[i,2]],

                    -K*(L-1)],

                {i,0,numstates-1}];

allinputs=Table[

                Take[Join[Table[0,{K}],

                        IntegerDigits[i,2]],

                    -K],

                {i,0,M-1}];

For[m=1,m<=M,m++,

 input=allinputs[[m]];

 For[i=1,i<=numstates,i++,

  state=allstates[[i]];

  output=Table[0,{n}];

  bigstate=Flatten[Prepend[state,input]];

  newstate=Take[bigstate,K*(L-1)];

  For[j=1,j<=n,j++,

   output[[j]]=

      Mod[Apply[Plus,bigstate*generators[[j]]],2];

      ];

  wi=Apply[Plus,input];

  wo=Apply[Plus,output];

  {{k}}=Position[allstates,newstate];
```

68

```
   If[k!=1,

      AppendTo[xsol[[k]],JJ NN^wi DD^wo x[i] ]];

   If[k==1 && i!=1,

      AppendTo[xsol[[ls]],JJ NN^wi DD^wo x[i] ]];

  ];

];

xsol=Apply[Plus,xsol,{1}];

solnset=Table[x[i]==xsol[[i]],{i,2,ls}];

elims=Table[x[i],{i,2,numstates}];

Cancel[(x[ls] /. Solve[solnset,x[ls],elims])/x[1]]

]
```

REFERENCES

Azizoglu, M. and Humblet, P. A., "Envelope Detection of Orthogonal Signals with Phase Noise," *Journal of Lightwave Technology*, Vol. 9, No. 10, pp. 1398-1410, October 1991.

DeLange, O. E., "Optical Heterodyne Detection," *IEEE Spectrum Magazine,* pp. 77-85, October 1968.

Foschini, G. J., Greenstein, L. J., and Vannucci, G., "Noncoherent Detection of Coherent Lightwave Signals Corrupted by Phase Noise," *IEEE Transactions on Communication*, Vol. 36, pp. 306-314, March 1988.

Foschini, G. J., and Vannucci, G., "Characterizing Filtered Light Waves Corrupted by Phase Noise," *IEEE Transactions on Information Theory*, Vol. 34, No. 6, pp. 1437-1448, November 1988.

Hamming, R. W., *Numerical Methods for Scientists and Engineers*, 2d ed., pp. 323-324, McGraw Hill, 1973.

Keiser, G., *Optical Fiber Communications,* 2d ed., McGraw Hill, 1991.

Larsen, K. J., "Short Convolutional Codes with Maximal Free Distance for Rates 1/2, 1/3, and 1/4," *IEEE Transactions on Information Theory*, Vol. IT-19, pp. 371-372, May 1973.

NASA Tech Briefs, *Add/Compare/Select Circuit for Rapid Decoding*, by J. M. Budinger, N. D. Becker, and P. N. Johnson, February 1993.

Odenwalder, J. P., *Optimal Decoding of Convolutional Codes*, Ph.D. Dissertation, University of California, Los Angeles, 1970.

Proakis, J. G., *Digital Communications,* 2d ed., McGraw Hill, 1989.

Salz, J., "Coherent Lightwave Communications," *AT&T Technical Journal,* Vol. 64, No. 10, pp. 2153-2209, December 1985.

Wolfram Research, Inc., *Mathematica*, Version 2.1, Wolfram Research, 1992

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center   2
   Cameron Station
   Alexandria, Virginia 22304-6145

2. Library, Code 52   2
   Naval Postgraduate School
   Monterey, California 93943-5002

3. Chairman, Code AA   1
   Department of Aeronautical and
   Astronautical Engineering
   Monterey, California 93943-5000

4. Chairman, Code EC   1
   Department of Electrical and
   Computer Engineering
   Monterey, California 93943-5000

5. Prof. T. T. Ha, Code EC/Ha   1
   Department of Electrical and
   Computer Engineering
   Monterey, California 93943-5000

6. Prof. D. J. Collins, Code AA/Co   1
   Department of Aeronautical and
   Astronautical Engineering
   Monterey, California 93943-5000

7. Prof. R. C. Robertson, Code EC/Rc   1
   Department of Electrical and
   Computer Engineering
   Monterey, California 93943-5000

8. Curricular Officer, Code 31   1
   Naval Postgraduate School
   Monterey, California 93943-5000

9. LT Thomas E. Neely   1
   USS John F. Kennedy (CV-67)
   FPO AE 09538-2800